```
                repeat_region      /rpt_family="MaLR"
                                   62712..62764
                                   /rpt_family="ID"
                repeat_region      65655..65793
                                   /rpt_family="Alu"
                repeat_region      65922..66103
                                   /rpt_family="MIR"
                repeat_region      67019..67319
                                   /rpt_family="MER121"
                repeat_region      67684..67898

Query Match        100.0%;  Score 25;  DB 10;  Length 178416;
Best Local Similarity  100.0%;  Pred. No. 1.2;
Matches   25;  Conservative  0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy         1 CAGTCCCAAATAAACAAACTGTCC 25
             |||||||||||||||||||||||||
Db     38485 CAGTCCCAAATAAACAAACTGTCC 38509


RESULT 4
AC101961
LOCUS       AC101961             162337 bp    DNA     linear   HTG 16-JUL-2003
DEFINITION  Mus musculus clone RP24-212A21, WORKING DRAFT SEQUENCE, 12
            unordered pieces.
ACCESSION   AC101961
VERSION     AC101961.3  GI:32813670
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 162337)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Mus musculus, clone RP24-212A21
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 162337)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
            Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
            Cook,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
            Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
            Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
            Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
            Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
            Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
            MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
            McCarthy,M., McEwan,P., McKernan,K., McPheeters,R., Meldrim,J.,
            Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
            Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
            Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
            Raymond,C., Retta,R., Rieback,M., Riley,M., Rise,C., Rogov,P.,
            Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
            Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
            Topham,K., Travis,M., Travis,N., Trigillio,J., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,N., Wu,X., Wyman,D., Ye,W.J., Young,G.,
            Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-NOV-2001) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 162337)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,M., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Cook,A., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,X., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-JUL-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Jul 16, 2003 this sequence version replaced gi:22381391.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ------------------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ------------------- Project Information
            Center project name: L17795
            Center clone name: 212_A21
            ------------------- Summary Statistics
            Sequencing vector: Plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 159608 bases at least Q40
            Consensus quality: 160658 bases at least Q30
            Consensus quality: 161096 bases at least Q20
            Insert size: 151000; agarose-fp
            Insert size: 161237; sum-of-contigs
            Quality coverage: 11.3 in Q20 bases; agarose-fp
            Quality coverage: 10.5 in Q20 bases; sum-of-contigs
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 12 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *     117780: contig of 117780 bp in length
            *     117781: gap of 100 bp
            *     117880: contig of 447 bp in length
            *     118327: gap of 100 bp
            *     118427: contig of 1032 bp in length
            *     118328: gap of 100 bp
            *     119101: contig of 674 bp in length
            *     119102: gap of 100 bp
            *     119201: contig of 752 bp in length
            *     119953: gap of 100 bp
            *     119954: contig of 100 bp
            *     120053: gap of 100 bp
            *     120054: contig of 1171 bp in length
            *     121224: gap of 100 bp
            *     121225: contig of 100 bp
            *     121324: gap of 100 bp
            *     121325: contig of 910 bp in length
            *     122234: gap of 100 bp
            *     122235: contig of 798 bp in length
            *     122335: gap of 100 bp
            *     123132: contig of 100 bp
            *     123232: gap of 100 bp
            *     123233: contig of 993 bp in length
            *     124225: gap of 100 bp
            *     124226: contig of 1032 bp in length
            *     125357: gap of 100 bp
            *     125358: contig of 1225 bp in length
            *     125458: gap of 100 bp
            *     126683: contig of 100 bp
            *     126782: gap of 100 bp
            *     126783: contig of 3200 bp in length
            *     129982: gap of 100 bp
            *     129983: contig of 100 bp
            *     130082: gap of 100 bp
            *     130083: contig of 32255 bp in length
                           Location/Qualifiers
            1..162337
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
FEATURES
     source
```

```
                   misc_feature    /clone="RP24-212A21"
                                   /clone_lib="RPCI-24 Male Mouse BAC"
                                   1..117780
                                   /note="assembly_fragment"
                                   clone_end:SP6
                                   vector_side:left"
                   misc_feature   117881..118327
                                   /note="assembly_fragment"
                   misc_feature   118428..119101
                                   /note="assembly_fragment"
                   misc_feature   119202..119953
                                   /note="assembly_fragment"
                   misc_feature   120054..121224
                                   /note="assembly_fragment"
                   misc_feature   121325..122234
                                   /note="assembly_fragment"
                   misc_feature   122335..123132
                                   /note="assembly_fragment"
                   misc_feature   123233..124225
                                   /note="assembly_fragment"
                   misc_feature   124326..125357
                                   /note="assembly_fragment"
                   misc_feature   125458..126682
                                   /note="assembly_fragment"
                   misc_feature   126783..129982
                                   /note="assembly_fragment"
                   misc_feature   130083..162337
                                   /note="assembly_fragment"

ORIGIN

Query Match      85.6%;    Score 21.4;   DB 2;   Length 162337;
Best Local Similarity    95.7%;    Pred. No. 44;
Matches 22;    Conservative    0;   Mismatches    1;   Indels    0;   Gaps    0;

Qy       1 CAGTCCACAAATAAACAAACTGT 23
           ||||||||||||||| |||||||
Db  135188 CAGTCCACAAATAAACAAACTGT 135210

RESULT 5
AL663052
LOCUS       AL663052             202972 bp    DNA    linear    ROD 05-APR-2002
DEFINITION  Mouse DNA sequence from clone RP23-202N14 on chromosome X, complete
            sequence.
ACCESSION   AL663052
VERSION     AL663052.8  GI:19572097
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS     Heath,P.
TITLE       Direct Submission
JOURNAL     Submitted (04-APR-2002) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Mar 21, 2002 this sequence version replaced gi:19168580.
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest. The following
            abbreviations are used to associate primary accession numbers given
            in the feature table with their source databases: Em:, EMBL; Sw:,
```

```
            SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information on the WORMPEP
            database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-202N14 is
            from the RPCI-23 Mouse PAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
            VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1..202972
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="X"
                     /clone="RP23-202N14"
                     /clone_lib="RPCI-23"

ORIGIN

Query Match      83.2%;    Score 20.8;   DB 10;   Length 202972;
Best Local Similarity    91.7%;    Pred. No. 79;
Matches 22;    Conservative    0;   Mismatches    2;   Indels    0;   Gaps    0;

Qy       2 AGTCCACAAATAAACAAACTGTCC 25
           |||||| ||||||||||||| ||||
Db    9211 AGTCCTCAAATAAACAAAGTGTCC 9234

RESULT 6
AC146500/c
LOCUS       AC146500             166829 bp    DNA    linear    HTG 21-AUG-2003
DEFINITION  Danio rerio clone CH211-130P22, *** SEQUENCING IN PROGRESS ***, 7
            ordered pieces.
ACCESSION   AC146500
VERSION     AC146500.1  GI:34013546
KEYWORDS    HTG; HTGS_PHASE2; HTGS_ACTIVEFIN.
SOURCE      Danio rerio (zebrafish)
ORGANISM    Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1 (bases 1 to 166829)
AUTHORS     Talbot,W.S., Rauch,G.J., Grimwood,J., Dickson,M., Schmutz,J. and
            Myers,R.M.
TITLE       Genomic studies of vertebrate diversity
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 166829)
AUTHORS     Talbot,W.S., Rauch,G.J., Grimwood,J., Dickson,M., Schmutz,J. and
            Myers,R.M.
TITLE       Direct Submission
JOURNAL     Submitted (21-AUG-2003) Stanford Human Genome Center, 975
            California Avenue, Palo Alto, CA 94304, USA
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 7 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * been provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *        9728: contig of 9728 bp in length
            * 9729   9828: gap of unknown length
            * 9829  10096: contig of 1068 bp in length
            * 10097 10996: gap of unknown length
            * 10097 75064: contig of 64068 bp in length
            * 75065 75164: gap of unknown length
            * 75165 120260: contig of 45096 bp in length
            * 120261 120360: gap of unknown length
            * 120361 130051: contig of 9691 bp in length
            * 130052 130151: gap of unknown length
            * 130152 152603: contig of 22452 bp in length
            * 152604 152703: gap of unknown length
            * 152704 166829: contig of 14126 bp in length
            The sequence of the clone was established as a mapping and
```

```
            sequencing collaboration at the Stanford Genome Evolution Center,
            funded by the NIH Centers of Excellence in Genomic Science (CEGS)
            initiative (http://cegs.stanford.edu). The clone was isolated from
            the BAC library CHORI-211 (http://bacpac.chori.org).
            ------Genome Center
                     Center: Stanford Human Genome Center
                     Center Code: shgc
                     Web site: http://www-shgc.stanford.edu
                     ------
                     Project Information
                     Center Project Name: 2367
                     Center clone name: CH211-130P22
                     ------
                     Summary Statistics
                     Consensus quality: 165931 bases at least Q40
                     Consensus quality: 166056 bases at least Q30
                     Consensus quality: 166078 bases at least Q20
                     Estimated insert size: 164000; agarose-fp estimation
                     Estimated insert size: 166122; sum-of-contigs
            estimation   Quality coverage:   1.01 in Q20 bases; agarose-fp
            estimation   Quality coverage:   1 in Q20 bases; sum-of-contigs
            estimation.
FEATURES      Location/Qualifiers
     source   1..166829
              /organism="Danio rerio"
              /mol_type="genomic DNA"
              /db_xref="taxon:7955"
              /clone="CH211-130P22"
ORIGIN

Query Match    81.6%;    Score 20.4;    DB 2;    Length 166829;
Best Local Similarity  95.5%;  Pred. No. 1.28e+02;
Matches  21;  Conservative  0; Mismatches  1;  Indels  0;  Gaps  0;

Qy      3 GTCCACAAATAAACAAACTGTC 24
          ||||||||||||| ||||||||
Db  96903 GTCCACAAATAAACATACTGTC 96882

RESULT 7
AL954819
LOCUS       AL954819    178491 bp    DNA    linear    HTG 30-OCT-2003
DEFINITION  Danio rerio clone CH211-1F7, WORKING DRAFT SEQUENCE, 2 unordered
            pieces.
ACCESSION   AL954819
VERSION     AL954819.17  GI:37606232
KEYWORDS    HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Danio rerio (zebrafish)
ORGANISM    Danio rerio
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
            Cypriniformes; Cyprinidae; Danio.
REFERENCE   1  (bases 1 to 178491)
AUTHORS     Pelan,S.
TITLE       Direct Submission
JOURNAL     Submitted (28-OCT-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Oct 9, 2003 this sequence version replaced gi:37496382.
            -------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: zfish-help@sanger.ac.uk
            -------------- Project Information
            Center project name: zC1F7
            -------------- Summary Statistics
            Assembly program: XGAP4; version 4.5
            Chemistry: Dye-terminator; 100% of reads
            Consensus quality: 178029 bases at least Q40
            Consensus quality: 178183 bases at least Q30
            Consensus quality: 178289 bases at least Q20
            Insert size: 178391; sum-of-contigs
            Insert size: 128908; 35.7% error; agarose-fp
            Quality coverage: 7.89x in Q20 bases; sum-of-contigs Quality
            coverage: 11.26x in Q20 bases; agarose-fp
            ----------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 2 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1  163111: contig of 163111 bp in length
            *   163112  163211: gap of 100 bp
            *   163212  178491: contig of 15280 bp in length.
FEATURES      Location/Qualifiers
     source   1..178491
              /organism="Danio rerio"
              /mol_type="genomic DNA"
              /db_xref="taxon:7955"
              /clone_lib="CHORI-211"
     misc_feature  1..163111
              /note="assembly_fragment:03239
              fragment_chain:1"
     misc_feature  163212..178491
              /note="assembly_fragment:03723
              fragment_chain:1
              clone_end:SP6
              vector_side:right"
ORIGIN

Query Match    80.8%;    Score 20.2;    DB 2;    Length 178491;
Best Local Similarity  88.0%;  Pred. No. 1.4e+02;
Matches  22;  Conservative  0; Mismatches  3;  Indels  0;  Gaps  0;

Qy      1 CAGTCCCACAAATAAACAAACTGTCC 25
          |||||||||||||||||| |||||||
Db  58234 CAGTCCCAAAGTAAACAAACTATCC 58258

RESULT 8
AC127100/c
LOCUS       AC127100    222745 bp    DNA    linear    HTG 11-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-164M9, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC127100
VERSION     AC127100.2  GI:23664869
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 222745)
AUTHORS     Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceaar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
```

```
            Consensus quality: 201828 bases at least Q20
            Estimated insert size: 205186; sum-of-contigs estimation
            Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
            ------------
            * NOTE: Estimated insert size may differ from sequence length
            *       (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *       consists of 2 contigs. The true order of the pieces
            *       is not known and their order in this sequence record is
            *       arbitrary. Gaps between the contigs are represented as
            *       runs of N, but the exact sizes of the gaps are unknown.
            *       This record will be updated with the finished sequence
            *       as soon as it is available and the accession number will
            *       be preserved.
            *       1      221310: contig of 221310 bp in length
            *       221311 221410: gap of unknown length
            *       221411 222745: contig of 1335 bp in length.
FEATURES             Location/Qualifiers
     source          1..222745
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-164M9"
     misc_feature    1..1603
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature    3684..4529
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH362192"
ORIGIN

Query Match              80.8%;  Score 20.2;  DB 2;  Length 222745;
Best Local Similarity    88.0%;  Pred. No. 1.4e+02;
Matches    22;  Conservative   0;  Mismatches   3;  Indels   0;  Gaps   0;

Qy        1 CAGTCCACAAATAAACAAACTGTCC 25
            |||||||||||||||||  ||||||
Db    68558 CAGCCCACAAATAAACAAACTCTTC 68534

RESULT 9
AX655462
LOCUS       AX655462                1019 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 5332 from Patent WO03000898.
ACCESSION   AX655462
VERSION     AX655462.1  GI:29158276
KEYWORDS
SOURCE      Oryza sativa
  ORGANISM  Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Chang,H.S., Chen,W., Cooper,B., Glazebrook,J., Goff,S.A., Hou,Y.M.,
            Katagiri,F., Quan,S., Tao,Y., Whitham,S., Xie,Z., Zhu,T. and Zou,G.
  TITLE     Plant genes involved in defense against pathogens
  JOURNAL   Patent: WO 03000898-A 5332 03-JAN-2003;
            Syngenta Participations AG (CH)
FEATURES             Location/Qualifiers
     source          1..1019
                     /organism="Oryza sativa"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:4530"
ORIGIN

Query Match              79.2%;  Score 19.8;  DB 6;  Length 1019;
Best Local Similarity    91.3%;  Pred. No. 3.7e+02;
Matches    21;  Conservative   0;  Mismatches   2;  Indels   0;  Gaps   0;

Qy        2 AGTCCACAAATAAACAAACTGTC 24
            ||||| ||||||||||| |||||
```

```
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,K., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Parks,K.,
            Nwaokelemen,O., Okwuonu,G., Olairnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,D., Wilson,R., Wieczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 222745)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 222745)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-OCT-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Oct 10, 2002 this sequence version replaced gi:21743892.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ----------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----------- Project Information
            Center project name: GNPW
            Center clone name: CH230-164M9
            ----------- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 198734 bases at least Q40
            Consensus quality: 200599 bases at least Q30
```

```
Db      657 AGTCCACAAATTAACAACCTGTC 679

RESULT 10
AF391109
LOCUS       AF391109    2993 bp    DNA    linear    PLN 02-JUL-2002
DEFINITION  Oryza sativa beta-expansin (EXPB16) gene, partial cds.
ACCESSION   AF391109
VERSION     AF391109.1  GI:21666631
KEYWORDS
SOURCE      Oryza sativa
ORGANISM    Oryza sativa
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1 (bases 1 to 2993)
AUTHORS     Lee,Y. and Kende,H.
TITLE       Expression of expansin genes in rice internodes
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 2993)
AUTHORS     Lee,Y. and Kende,H.
TITLE       Direct Submission
JOURNAL     Submitted (19-OCT-2001) MSU-DOE Plant Research Laboratory, Michigan
            State University, East Lansing, MI 48824, USA
FEATURES             Location/Qualifiers
     source          1..2993
                     /organism="Oryza sativa"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:4530"
                     /note="this sequence is from the Monsanto rice genome
                     database; www.rice-research.org; contig OSM26829"
     gene            <2388..>2993
                     /gene="EXPB16"
     mRNA            join(<2388..2553,2663..2775,2916..>2993)
                     /gene="EXPB16"
                     /product="beta-expansin"
     CDS             join(2388..2553,2663..2775,2916..>2993)
                     /gene="EXPB16"
                     /codon_start=1
                     /product="beta-expansin"
                     /protein_id="AAM73780.1"
                     /db_xref="GI:21666632"
                     /translation="MALAAKLLPSIVAFVALACCVLRSSVASVDHHRKLSGWSIGGAT
                     WYGPANGSGTDGGACGYQGDVGGQPPFNSMIAAGSPSIYESGKGCGSCYQVKCSGNPSC
                     SGKPVTVLTDLCPGGA"

ORIGIN

Query Match     79.2%;  Score 19.8;  DB 8;  Length 2993;
Best Local Similarity   91.3%;  Pred. No. 3.3e+02;
Matches    21;  Conservative   0;  Mismatches  2;  Indels   0;  Gaps   0;

Qy        2 AGTCCACAAATAAACAACCTGTC 24
            |||||||||| ||||||||||||
Db     2025 AGTCCACAAATTAACAACCTGTC 2047

RESULT 11
AC116314
LOCUS       AC116314    26845 bp    DNA    linear    HTG 26-MAR-2002
DEFINITION  Trypanosoma cruzi chromosome 3 clone chimeric cosmid Tcc44h21-2
            strain CL Brener. *** SEQUENCING IN PROGRESS ***.
ACCESSION   AC116314
VERSION     AC116314.1  GI:19718715
KEYWORDS    HTG; HTGS_PHASE1.
SOURCE      Trypanosoma cruzi
ORGANISM    Trypanosoma cruzi
            Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
            Trypanosoma; Schizotrypanum.
REFERENCE   1 (bases 1 to 26845)
AUTHORS     Andersson,B. and Bontempi,E.J.
JOURNAL     Unpublished
REFERENCE   2 (bases 1 to 26845)
AUTHORS     Andersson,B. and Bontempi,E.J.
TITLE       Direct Submission
JOURNAL     Submitted (26-MAR-2002) Department of Genetics and Pathology,
            Uppsala University, Rudbeck Laboratory, Uppsala SE-751 85, Sweden
COMMENT     * NOTE: This is a 'working draft' sequence. It currently
            * consists of 1 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            * 1    26845: contig of 26845 bp in length.

FEATURES             Location/Qualifiers
     source          1..26845
                     /organism="Trypanosoma cruzi"
                     /mol_type="genomic DNA"
                     /strain="CL Brener"
                     /db_xref="taxon:5693"
                     /chromosome="3"
                     /clone="chimeric cosmid Tcc44h21-2"
     gene            complement(1495..2742)
                     /gene="Tcc44h21-2.1"
     CDS             complement(1495..2742)
                     /gene="Tcc44h21-2.1"
                     /note="Predicted by Glimmer and TestCode.
                     Motifs: PF01066 CDP-alcohol phosphatidyltransferase;
                     PS00379 CDP_ALCOHOL_P_TRANSF.
                     8 transmembrane domains predicted by SOSUI and TMHMM2"
                     /codon_start=1
                     /product="Tcc44h21-2.1"
                     /protein_id="AAL96364.1"
                     /db_xref="GI:19718716"
                     /translation="MDLFLSVSPTDDRPLVPPQLSYLTEHVLTPFYDLVVLYYPASWT
                     PNKVTLFGIFMTLLSSLLLTGMPPTLFEPPYATIRPASFLGEDSKWQNFFGPTPLH
                     PTTLRPLWSSIFPSGNWLLFTCGILNSIYCIADNTDGRLAARLKKSSCIGEYLDHGLD
                     CVTSLLSTCVALSVFGISFSNISLAVVTVAMATVFSHTLNYEKHIFIWGNRFVSVDEA
                     MIFFCVVHWIPLLFPGLASARVSPALLYAVLPEKWAHALLPLRYVEAIMVIYWASQVY
                     VLIDIALNQWKMIFRTTLALLLNTVLMLGVIPYQTAQLVKVDPAGVWGFSYVAIW
                     FITVSCTSSTVVHILIYSHCARLPVDPTALAGLLFVWLAFINCPAAGMMISVVWHIA
                     QILCYVDGLQNRHLEGSQKVKVV"
     gene            complement(3458..3874)
                     /gene="Tcc44h21-2.2"
     CDS             complement(3458..3874)
                     /gene="Tcc44h21-2.2"
                     /note="Predicted by Glimmer and TestCode"
                     /codon_start=1
                     /product="Tcc44h21-2.2"
                     /protein_id="AAL96365.1"
                     /db_xref="GI:19718717"
                     /translation="MLRRTIVPLAKVYGAAPNGTGCKQPVRHRLLHGKRERQGSLFA
                     IANDVKYDEKRLRQQLNAMLBERLPPRTRLBHNKANGGEALRSVGLSTCRVWSDAVT
                     CPPIRSQGCSFSTKGIMHCTTAHMTTPVYRTKIAYK"
     gene            complement(4307..5155)
                     /gene="Tcc44h21-2.3"
     CDS             complement(4307..5155)
                     /gene="Tcc44h21-2.3"
                     /note="Predicted by Glimmer (2nd Met) and TestCode.
                     BlastP: 31% identical to gi|19114099|ref|NP_593187.1|
                     (NC_003424) dna-directed rna polymerase i ii and iii 24 kd
                     polypeptide [Schizosaccharomyces pombe]; 27% to
                     gi|13812223|ref|NP_113354.1| (NC_002752) RNA polymerase I,
                     II and III 24.3 kDa subunit [Guillardia theta]; 25% to
                     gi|15228724|ref|NP_188871.1| (NM_113130) RNA polymerase I,
                     II and III 24.3 kDa subunit [Arabidopsis thaliana]; etc.
                     Motifs: PF01191 RNA polymerases H / 23 kDa subunit.
                     EST hits: gi|482741|gb|AI668107.1|AI668107 TENG0983 T.
                     cruzi epimastigote normalised cDNA Library Trypanosoma
                     cruzi cDNA clone n582.r 5;
                     gi|3017817|gb|AA890938.1|AA890938  TENS0788 T. cruzi
                     epimastigote normalized cDNA Library Trypanosoma cruzi
                     cDNA clone 788 5'"
                     /codon_start=1
                     /product="Tcc44h21-2.3"
```

```
                AGIQRGDQIWEVGRHVRTLKDFRKILASTDGTHICVGIRKKSSGALTRIRLRRGWRG
                PLESEGSPSTNSSPLEPRIISMLGKAGSAFLASAMNRAHSAFVFPKSVQQSYLPLNNTE
                HGS"
gene            complement(11826. .12257)
                /gene="Tcc44h21-2.6"
CDS             complement(11826. .12257)
                /gene="Tcc44h21-2.6"
                /note="Predicted by Glimmer and TestCode.
                Motifs: PS00028 ZINC_FINGER_C2H2_1; SM0451 Family of
                C2H2-type zinc fingers, present in matrin, U1 small
                nuclear ribonucleoprotein C and other RNA-binding
                proteins"
                /codon_start=1
                /product="Tcc44h21-2.6"
                /protein_id="AAL96369.1"
                /db_xref="GI:19718721"
                /translation="MCLFPPCFAFRIFDRQRRVFFVQEGGWSAAVSTDIIIITTTVV
                ASFSQDVMFHTPVCGRSAGAFYCPSCNVYCSDSRTALHRSSLKHKKKSGELEMERQL
                YKEDASVTVEDVMALVERKRVELGVVPWSQLRFTEEBTHAD"
gene            complement(12687. .16391)
                /gene="Tcc44h21-2.7"
CDS             complement(12687. .16391)
                /gene="Tcc44h21-2.7"
                /note="Predicted by Glimmer and TestCode"
                /codon_start=1
                /product="Tcc44h21-2.7"
                /protein_id="AAL96370.1"
                /db_xref="GI:19718722"
                /translation="MNTPTKLGDLVIATPSLSAAGKSCAAEVGTTATSRSRALVVQHK
                KETAKATVKEVDVLHRGLKRLETRMHELESRVQYVNPWVVQYMWELLCELRDRIGMNY
                SHOREKAWITGLRRRRENSRGDQPLTESGMKTPLKQGPTEARTGTAGSRGETLLTTPT
                HRFEYSNEGVEGEQQEDDSILDTSVSDESDDGRLPFRENGVVTTEACAQGADFVQS
                SPIDDSKRLKVGRTPKTDGRRLARHMATLGARWEVVRITQBSTSRIEIALEESTQR
                SLLISSLSCSLATVLHQREVVLASSASSNGNVHGMRNATQRQGGIATQRRRERFPD
                LDVPSPPRTEIPVGLSPIQLAQSPSDFFSPERGNGRGGNQERSRRAVRSATDVVDDD
                SDFFSPNNRQKSRREHEMTALACSVLEAPEFMGRVNEVSRQLAAQIQQTDMSGEKH
                LSSQSLPSALVLTAREAVNAIAGAHDEALYHQLREYLAHEQRQRTTKEESLWQRVTE
                YEEMWERRTAAERQTKEEVLQEHQRLLALGREDAESCSEIRRELLKLCGEVRAVEGM
                PDTFAASRPSTELDDRQNGTDAIRDEKVFLRLETLSAAVSSLYSDLESLRNKSKGEL


Query Match    79.2%;   Score 19.8;  DB 2;  Length 26845;
Best Local Similarity    91.3%;  Pred. No. 2.6e+02;
Matches    21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    3  GTCCACAAATAAACAAACTGTCC 25
         |||||||| ||| ||||||| ||
Db  8821 GTCCACAAAAAAACAAACTTTCC 8843


RESULT 12
AC135183/c
LOCUS       AC135183           86992 bp    DNA        linear   HTG 27-MAR-2003
DEFINITION  Homo sapiens chromosome 17 clone CTD-2182D19 map 17, 2 ordered
            pieces.
ACCESSION   AC135183
VERSION     AC135183.7  GI:27545099
KEYWORDS    HTG; HTGS_PHASE2; HTGS_FULLTOP; HTGS_CANCELLED.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 86992)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone CTD-2182D19
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 86992)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
```

```
                /protein_id="AAL96366.1"
                /db_xref="GI:19718718"
                /translation="MGSIELKIVRLYRAYNTAIQMCMDRGYTVRHPSAVAQAVRDPC
                VYSYEHGLDYDWFLQRFVTQRAAASGAERAEEDEDDKVSPTNQNKLLEENKRILMR
                GAMRLVCSAELDGATDDEVSLSKRGEKKGSLMLVFFSASPSLSMSELQEFREKALLK
                EARCMVVVSAGKIFPMVRRGARELQGFLRAGTGEIMSLQLLBEDELAYNIARHDTVP
                QHEVLSPEEARLFLQQRQISLSQLPRILKEDPMVAYLGLARGSIVKIHRDTKESGPYE
                MYRQVI"
gene            complement(5436. .6095)
                /gene="Tcc44h21-2.4"
CDS             complement(5436. .6095)
                /gene="Tcc44h21-2.4"
                /note="Predicted by Glimmer and TestCode.
                BlastP: 38% identical to gi|17507947|ref|NP_491961.1|
                (NM_059560) RNA polymerase [Caenorhabditis elegans]
                gi|7206747|gb|AAF39906.1| (AC006665) contains similarity
                to Pfam family PF01191 (RNA polymerases H / 23 Kd
                subunit), score=163.3, E=4.1e-45, N=1 [Caenorhabditis
                elegans]; 37% togi|1680605|gb|AAC60556.1|(S59774) (S59774)
                RNA polymerase subunit [Saccharomyces cerevisiae] ; 37% to
                gi|12229927|sp|Q9P4B9|RPB5_KLUMA DNA-directed RNA
                polymerases II 24 kDa polypeptide (RNA polymerase II
                subunit 5) gi|9651716|gb|AAF91238.1|AF25206_4 (AF225206)
                RNA polymerase subunit [Kluyveromyces marxianus]; etc.
                Motifs: PF01191 RNA polymerases H / 23 kDa subunit;
                PS01110 RNA polymerases H / 23 Kd subunits signature"
                /codon_start=1
                /product="Tcc44h21-2.4"
                /protein_id="AAL96367.1"
                /db_xref="GI:19718719"
                /translation="MGDESNAIGQEALLYDKMFRVMRTVGEMMCDRKYQVADKIIPSS
                VEEFKDWYTDPREGVIYRDRMTVPCERIGVSQRSRAMVFFSEGLSMETMKRYHTQASE
                ANCDRVIIVVQGRINAIVKRHVEDINRSSLGLKIQLFDEDDLVVNITHHELVPKHTQL
                EDEEVKEMLQAHALELSALPRILSTDPVAAYLGLERGRVRIERKSMSAGFVVTYRQV
                V"
gene            complement(6881. .11548)
                /gene="Tcc44h21-2.5"
CDS             complement(6881. .11548)
                /gene="Tcc44h21-2.5"
                /note="Predicted by Glimmer and TestCode. Also called DHR (Dlg
                homologous region) or GLGF (relatively well conserved
                tetrapeptide in these domains). They may function in
                targeting signalling molecules to sub-membranous sites"
                /codon_start=1
                /product="Tcc44h21-2.5"
                /protein_id="AAL96368.1"
                /db_xref="GI:19718720"
                /translation="MGEGRCESFPHLETDLPSRNEACQNAVGVLEQILSISDLHMAAR
                ATLHETQCSSVPTLSSMTETIAHNSQLVCERTVALLKKLVPEPREVSEGEVAGGROLS
                SPKIADLPSFVQKSRLLHERVQRVAVTLGGLKHSRLLGEESAGRRIITERASLAIKQL
                LEVLLAVEA'VDETPLSSRAGQGTARSRYFVLDSQRSSEPLEAPSTESSHAASEMLFS
                GEINKDNPPKSISFENGSQGKKINRNSPRYLTTTRLRPTAEKREEREGSKTRTSCCEG
                KHLPFLQDERKESSSNSFYSCFSASVSSGNADTMGQSSSLRPYDKKGASPTSPKFREI
                NCGNADSLETIASTRCIPLEELCSINTQLLAFKNKLLPNTTVRVPIRNEDLSTAVPT
                RVNVHRRIFPQGKLWRELLCNEDPTTFSRWKELFIEEVVLLFGIPHRWVSDVQLANDTQ
                TTMTGDQAVKFRVRLPLSLEANEVNTRIQNHEFISMLLYEELQKKEEEGIKEISQ
                SDTNSFLSVASRDNPBSMSPPEFANTAVPNSDAGSSDSRSGSGFDDCHNCLSNKIKND
                KLLLDASSAIIQKEIAIGKEILSRDGLIIVQEKYTRKLVMKKSTONQHMPGHEQCQGKKPVDVSL
                PTYRNHTGQQEPQIEASDALEEVSEFRFPNTHPACVLPLPPQSFAFPRHEHVSASTV
                DEEKTELSASSPSPVRTDCEVNDADYKEKSIFGNGSLLMNPHLPESSEREFVLGPR
                KAWEPTVQWFYTSHEKVISGERWDLVLEKRDELRRTFIAEVAALFELPEENVTAVEF
                VLGSLHATFNLMHDRYLKAAEIDALLTVFDFPQVRILYHRCVEGVNAFPPVSNKMSG
                SGYERQNVNANDLRMQLPFDNSIAGKNTSSFRNECADEKLVRLASEESLFFCGRQLFP
                VRLSMNSLQENAATPDNEPITFAHLALLHDISTDALVRANPHLEAFGVNDPLPSSAHI
                AREMVSNGQALRQRSTHFHSDGNLDPISRNCPVNVSVGSLEASDSPVFLPAACASYQS
                SGTLIGMREESPTAVVVSGSDGEKKTVGFGLESTPFKQLTQESADGNGHALATGSEPK
                GVHKKWHDFASLHQVKQHLLAPPQRSRPRMHVSDLRVKPSFSTIRDAVRRDQAVL
                LMLRSTEQFILLRFFARWQYVARIRKGHCEYLALQRVPLEGSPTKSPIQETESLNSRF
                REKTLGGTLTPQHVVPSKTRSLRGQNAELIAKTIBAIARTEKRGVSSRNAHQNIPVQK
                KSFFPAQQNMVVVSSTHAISVSAPMRTFGGDHSSQSGYFAHTGSRESSRSPSSCSS
                SPHLAFSGMSSDHSGYSASAFVHRPNVVPDLGISERPPLGLCVSSLLIVTEVHGEAEA
```

Best Local Similarity 91.3%; Pred. No. 2.3e+02;
Matches 21; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```
Qy     1 CAGTCCACAATAAACAAACTGT 23
           ||||  ||||||||||||||
Db 61466 CAGTGCCAAATAAACAAACTGT 61444
```

RESULT 13
HUMNEUROF
LOCUS   HUMNEUROF   100849 bp   DNA   linear   PRI 20-SEP-1995
DEFINITION   Human oligodendrocyte myelin glycoprotein (OMG) exons 1-2; neurofibromatosis 1 (NF1) exons 28-49; ecotropic viral integration site 2B (EVI2B) exons 1-2; ecotropic viral integration site 2A (EVI2A) exons 1-2; adenylate kinase (AK3) exons 1-2.
ACCESSION   L05367 L03723
VERSION   L05367.1 GI:189152
KEYWORDS   adenylate kinase; ecotropic viral integration site 2A; ecotropic viral integration site 2B; neurofibromatosis type 1; oligodendrocyte myelin glycoprotein.
SOURCE   Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 100849)
  AUTHORS   Cawthon,R.M., Weiss,R.B., Xu,G., Viskochil,D., Culver,M., Stevens,J., Robertson,M., Dunn,D., Gesteland,R., O'Connell,P. and White,R.
  TITLE   A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations
  JOURNAL   Cell 62 (1), 193-201 (1990)
  MEDLINE   90304909
   PUBMED   2114220
REFERENCE   2 (bases 1 to 100849)
  AUTHORS   Cawthon,R.M., O'Connell,P., Buchberg,A.M., Viskochil,D., Weiss,R.B., Culver,M., Stevens,J., Jenkins,N.A., Copeland,N.G. and White,R.
  TITLE   Identification and characterization of transcripts from the neurofibromatosis 1 region: the sequence and genomic structure of EVI2 and mapping of other transcripts
  JOURNAL   Genomics 7 (4), 555-565 (1990)
  MEDLINE   90053953
   PUBMED   2117566
REFERENCE   3 (bases 1 to 100849)
  AUTHORS   Cawthon,R.M., Andersen,L.B., Buchberg,A.M., Xu,G.F., O'Connell,P., Viskochil,D., Weiss,R.B., Wallace,M.R., Marchuk,D.A., Culver,M.Stevens,J., Jenkins,N.A., Copeland,N.G., Collins,F.S. and White,R.
  TITLE   cDNA sequence and genomic structure of EVI2B, a gene lying within an intron of the neurofibromatosis type 1 gene
  JOURNAL   Genomics 9 (3), 446-460 (1991)
  MEDLINE   91236164
   PUBMED   1903357
REFERENCE   4 (bases 1 to 100849)
  AUTHORS   Viskochil,D., Cawthon,R., O'Connell,P., Xu,G.F., Stevens,J., Culver,M., Carey,J. and White,R.
  TITLE   The gene encoding the oligodendrocyte-myelin glycoprotein is embedded within the neurofibromatosis type 1 gene
  JOURNAL   Mol. Cell. Biol. 11 (2), 906-912 (1991)
  MEDLINE   91117257
   PUBMED   1899288
REFERENCE   5 (bases 1 to 100849)
  AUTHORS   Wallace,M.R., Andersen,L.L., Letcher,R., Odeh,H., Saulino,A.M., Fountain,J., Brereton,A., Nicholson,J., Mitchell,A., Brownstein,B.H. and Collins,F.
  TITLE   A de novo Alu insertion results in neurofibromatosis type 1
  JOURNAL   Nature 353 (6347), 864-866 (1991)
  MEDLINE   92049738
   PUBMED   1719426
REFERENCE   6 (bases 1 to 100849)
  AUTHORS   Weiss,R.B., Dunn,D., DiSera,L., Wheatley,W., Kimball,A., Rote,C., Cherry,J., Duval,B., Lee,R., Ferguson,M.W.J. and Gesteland,R.F.
  TITLE   The Human Neurofibromatosis Type 1 locus: genomic sequence of the

             Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu.G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (08-OCT-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3 (bases 1 to 86992)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Bookhgalter,B., Canarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A., Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N., Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H., Viel,R., Vo.A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE   Direct Submission
  JOURNAL   Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Jan 8, 2003 this sequence version replaced gi:27372474.
             All repeats were identified using RepeatMasker:
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
             ------- Genome Center
             Center: Whitehead Institute/ MIT Center for Genome Research
             Center code: WIBR
             Web site: http://www-seq.wi.mit.edu
             Contact: sequence_submissions@genome.wi.mit.edu
             ------- Project Information
             Center project name: L28311
             Center clone name: 2182_D_19
             ----
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 2 contigs. Gaps between the contigs
           * are represented as runs of N. The order of the pieces
           * is believed to be correct as given, however the sizes
           * of the gaps between them are based on estimates that have
           * provided by the submittor.
           * This sequence will be replaced
           * by the finished sequence as soon as it is available and
           * the accession number will be preserved.
           *   1   47027: contig of 47027 bp in length
           * 47028   47127: gap of 100 bp
           * 47128   86992: contig of 39865 bp in length.
FEATURES   Location/Qualifiers
     source   1..86992
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="CTD-2182D19"
                     /clone_lib="CITD1 Human BAC"
ORIGIN

Query Match   79.2%; Score 19.8; DB 2; Length 86992;

```
JOURNAL      3' region
             Unpublished (1992)
REFERENCE    7 (bases 1 to 100849)
AUTHORS      Xu,G., O'Connell,P., Stevens,J. and White,R.
TITLE        Characterization of human adenylate kinase 3 (AK3) cDNA and mapping
             of the AK3 pseudogene to an intron of the NF1 gene
JOURNAL      Genomics 13 (3), 537-542 (1992)
MEDLINE      92347846
PUBMED       1639383
COMMENT      Original source text: Homo sapiens DNA.
             Submitting author Robert Weiss may be contacted at the following
             address and telephone number:                     e-mail:
             bob@corona.med.utah.edu                            postal: Room
             7160, Eccles Institute of Human Genetics, University of Utah, Salt
             Lake City, Utah 84112     telephone: (801) 585-3436.
FEATURES             Location/Qualifiers
     source          1..100849
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /map="17q11.2"
     intron          <1..16919
                     /number=1
     gene            complement(5761..8342)
                     /gene="OMG"
     exon            complement(5761..7447)
                     /gene="OMG"
                     /note="G00-127-563"
                     /citation=[4]
                     /number=1
     intron          complement(7448..6265)
                     /gene="OMG"
                     /note="G00-127-563"
                     /number=2
     exon            complement(8266..8342)
                     /gene="OMG"
                     /note="G00-127-563"
                     /citation=[4]
                     /number=1
     repeat_region   12224..12233
                     /rpt_family="GT"
     gene            complement(14658..25133)
                     /gene="EVI2B"
     exon            complement(14658..16737)
                     /gene="EVI2B"
                     /note="G00-125-901"
                     /citation=[3]
                     /number=2
     intron          complement(16738..25076)
                     /gene="EVI2B"
                     /note="G00-125-901"
                     /number=1
     repeat_region   24830..24856
                     /rpt_family="GT"
     exon            complement(25077..25133)
                     /gene="EVI2B"
                     /note="G00-125-901"
                     /citation=[3]
                     /number=1
     repeat_region   28465..28503
                     /rpt_family="AC"
     gene            complement(28542..32788)
                     /gene="EVI2A"
     exon            complement(28542..30121)
                     /gene="EVI2A"
                     /note="G00-125-191"
                     /citation=[2]
                     /number=2
     intron          complement(30122..32588)
                     /gene="EVI2A"
                     /note="G00-125-191"
                     /number=1
     exon            complement(32589..32788)
```

```
                     /gene="EVI2A"
                     /note="G00-125-191"
                     /citation=[2]
                     /number=1
     repeat_region   33193..33232
                     /rpt_family="AC"
     gene            36920..86010
                     /gene="NF1"
     exon            36920..37340
                     /gene="NF1"
                     /note="36920 = nt. 5032 from full-length neurofibromin
                     cDNA (HUMNF1B,M89914); G00-120-231"
                     /number=28
     intron          37341..38596
                     /gene="NF1"
                     /note="G00-120-231"
                     /number=28
     exon            38597..38937
                     /gene="NF1"
                     /note="G00-120-231"
                     /number=29
     intron          38938..41643
                     /gene="NF1"
                     /note="G00-120-231"
                     /number=29
     exon            41644..41846
                     /gene="NF1"
                     /note="G00-120-231"
                     /number=30
     intron          41847..46170
                     /gene="NF1"
                     /note="G00-120-231"
                     /number=30
     exon            46171..46364
                     /gene="NF1"
                     /note="G00-120-231"
                     /number=31
     variation       46216
                     /gene="NF1"
                     /note="leu to pro change, exon 31; G00-120-231"
                     /citation=[1]
                     /replace="c"
     variation       46260
                     /gene="NF1"
                     /note="arg to stop change, exon 31; G00-120-231"
                     /citation=[1]
                     /replace="t"
     intron          46365..47772
                     /gene="NF1"
                     /note="G00-120-231"
                     /number=31
     variation       46600..46740
                     /gene="NF1"
                     /note="t17;22 translocation breakpoint; G00-120-231"
                     /replace=""
     exon            47773..47913
                     /gene="NF1"
                     /note="G00-120-231"
                     /number=32
     intron          47914..48075
                     /gene="NF1"
                     /note="G00-120-231"
                     /number=32
     variation       48030

Query Match         79.2%;  Score 19.8;  DB 9;  Length 100849;
Best Local Similarity  91.3%;  Pred. No. 2.3e+02;
Matches   21;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy      1 CAGTCCACAAATAAACAAACTGT 23
          ||||| |||||||||||||||||
Db  32106 CAGTGCCCAAATAAACAAACTGT 32128
```

```
RESULT 14
AC134669
LOCUS       AC134669              113200 bp    DNA     linear   PRI 02-FEB-2003
DEFINITION  Homo sapiens chromosome 17, clone CTD-2370N5, complete sequence.
ACCESSION   AC134669
VERSION     AC134669.4  GI:28195535
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 113200)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 17, clone CTD-2370N5
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 113200)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
            Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
            Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,
            Liu,G., MacLean,C., Macdonald,P., Major,J., Matthews,C.,
            McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
            Murphy,T., Naylor,J., Nguyen,J., Nicol,R., Norbu,C., Norman,C.H.,
            O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K.,
            Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
            Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
            Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
            Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
            Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (28-SEP-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 113200)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,J., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-JAN-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 113200)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abcuelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,S., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R., MacLean,C.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-FEB-2003) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Feb 2, 2003 this sequence version replaced gi:27753749.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            --------- Genome Center
              Center: Whitehead Institute/ MIT Center for Genome Research
              Center code: WIBR
              Web site: http://www-seq.wi.mit.edu
              Contact: sequence_submissions@genome.wi.mit.edu
            --------- Project Information
              Center project name: L28272
              Center clone name: 2370_N_5
            ---------
FEATURES             Location/Qualifiers
     source          1..113200
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="17"
                     /map="17"
                     /clone="CTD-2370N5"
                     /clone_lib="CITB1 Human BAC"
     repeat_region   22..327
                     /rpt_family="AluSq"
     repeat_region   1062..1082
                     /rpt_family="AT_rich"
     repeat_region   complement(1195..1275)
                     /rpt_family="MIR"
     repeat_region   1962..2079
                     /rpt_family="FLAM_C"
     repeat_region   2081..2221
                     /rpt_family="(TAGA)n"
     repeat_region   2223..2508
                     /rpt_family="AluJo"
     repeat_region   2516..2572
                     /rpt_family="(TAAAAA)n"
     repeat_region   2696..2783
                     /rpt_family="MIR"
     repeat_region   2868..2890
                     /rpt_family="AT_rich"
     repeat_region   complement(2937..3009)
                     /rpt_family="HALib"
     repeat_region   3177..3486
                     /rpt_family="AluSq"
     repeat_region   3501..3755
                     /rpt_family="AluJo"
     repeat_region   complement(4077..4166)
                     /rpt_family="L1MEd"
     repeat_region   5144..5216
                     /rpt_family="L1ME"
     repeat_region   complement(5996..6291)
                     /rpt_family="MIR"
     repeat_region   6761..6973
                     /rpt_family="AluSq"
     repeat_region   7407..7702
                     /rpt_family="L2"
     repeat_region   complement(7758..8065)
                     /rpt_family="AluSc"
     repeat_region   8422..8536
                     /rpt_family="AluJo"
     repeat_region   complement(8551..9116)
                     /rpt_family="L1MC/D"
                     /rpt_family="MLTIk"
```

```
VERSION     AP005289.1  GI:21280339
KEYWORDS    HTG; HTGS_PHASE2.
SOURCE      Oryza sativa (japonica cultivar-group)
  ORGANISM  Oryza sativa (japonica cultivar-group)
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
            Ehrhartoideae; Oryzeae; Oryza.
REFERENCE   1
  AUTHORS   Sasaki,T., Matsumoto,T. and Katayose,Y.
  TITLE     Oryza sativa nipponbare(GA3) genomic DNA, chromosome 2, BAC
            clone:OJ1112_F09
  JOURNAL   Published Only in Database (2002)
REFERENCE   2  (bases 1 to 139503)
  AUTHORS   Sasaki,T., Matsumoto,T. and Katayose,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (29-MAY-2002) Takuji Sasaki, National Institute of
            Agrobiological Sciences, Rice Genome Research Program; Kannondai
            2-1-2, Tsukuba, Ibaraki 305-8602, Japan
            (E-mail:tsasaki@nias.affrc.go.jp, URL:http://rgp.dna.affrc.go.jp/,
            Tel:81-298-38-7441, Fax:81-298-38-7468)
COMMENT     The nucleotide sequence of this BAC clone was generated by
            combining Monsanto and RGP-Japan sequencing data.
            NOTE: It currently consists of 1 contigs. Gaps between the contigs
            are represented as runs of N. The order of the pieces is believed
            to be correct as given, however the sizes of the gaps between them
            are based on estimates that have provided by the submitter. This
            sequence will be replaced by the finished sequence as soon as it is
            available and the accession number will be preserved.
            * NOTE: This is a 'working draft' sequence.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
FEATURES             Location/Qualifiers
     source          1..139503
                     /organism="Oryza sativa (japonica cultivar-group)"
                     /mol_type="genomic DNA"
                     /cultivar="Nipponbare"
                     /db_xref="taxon:39947"
                     /chromosome="2"
                     /clone="OJ1112_F09"
ORIGIN
Query Match          79.2%;  Score 19.8;  DB 2;  Length 139503;
Best Local Similarity   91.3%;  Pred. No. 2.2e+02;
Matches    21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      2 AGTCCACAAATAAACAAACTGTC 24
          ||||||||||||| |||| |||||
Db  90916 AGTCCACAAATTAACAACCTGTC 90938

Search completed: March 25, 2004, 12:49:54
Job time : 175.338 secs
```

```
repeat_region   9387..9701
                /rpt_family="AluD"
repeat_region   complement(9785..9880)
                /rpt_family="MIR"
repeat_region   complement(9911..10069)
                /rpt_family="MER5A"
repeat_region   9930..10070
                /rpt_family="MER5A"
repeat_region   complement(10071..10134)
                /rpt_family="MIR"
repeat_region   10510..10530
                /rpt_family="MIR"
repeat_region   10533..10837
                /rpt_family="AT_rich"
repeat_region   11836..11856
                /rpt_family="AluSq"
repeat_region   12376..12401
                /rpt_family="AT_rich"
repeat_region   complement(12407..12550)
                /rpt_family="(CAA)n"
repeat_region   14242..14347
                /rpt_family="MIR3"
repeat_region   14349..14715
                /rpt_family="Tigger7"
repeat_region   complement(14...)
                /rpt_family="Tigger7"
repeat_region   14996..15284
                /rpt_family="AluY"
repeat_region   15510..15758
                /rpt_family="MER46C"
repeat_region   15785..15892
                /rpt_family="5S"
repeat_region   complement(15893..16321)
                /rpt_family="MLT1J2"
repeat_region   16322..16548
                /rpt_family="Tigger2a"
repeat_region   complement(...)
                /rpt_family="MLT1J2"
repeat_region   16555..16820
                /rpt_family="LTR33"
repeat_region   16949..16997
                /rpt_family="LTR33"
repeat_region   17026..17181
                /rpt_family="FRAM"
repeat_region   17427..17479
                /rpt_family="T-rich"
repeat_region   17544..17571
                /rpt_family="(T)n"
repeat_region   complement(17645..17954)
                /rpt_family="AluSq"
repeat_region   18006..18297
                /rpt_family="AluJo"
repeat_region   18351..18392
                /rpt_family="5S"
repeat_region   18540..18653
                /rpt_family="FLAM_A"
repeat_region   complement(19172..19286)
                /rpt_family="MER58C"
repeat_region   19287..19580
                /rpt_family="AluSg"
repeat_region   complement(19581..19676)

Query Match          79.2%;  Score 19.8;  DB 9;  Length 113200;
Best Local Similarity   91.3%;  Pred. No. 2.3e+02;
Matches    21;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy      1 CAGTCCACAAATAAACAAACTGT 23
          |||| |||||||||||||||||||
Db  60048 CAGTGCCCAAATAAACAAACTGT 60070

RESULT 15
AP005289
LOCUS       AP005289      139503 bp    DNA     linear    HTG 31-MAY-2002
DEFINITION  Oryza sativa (japonica cultivar-group) chromosome 2 clone
            OJ1112_F09, *** SEQUENCING IN PROGRESS ***.
ACCESSION   AP005289
```

OM nucleic - nucleic search, using sw model

```
                 GenCore version 5.1.6
            Copyright (c) 1993 - 2004 Compugen Ltd.


Run on:        March 25, 2004, 09:53:09 ; Search time 188.926 Seconds
                                           (without alignments)
                                           3951.570 Million cell updates/sec

Title:         US-09-963-285-1_COPY_378_402
Perfect score: 25
Sequence:      1 cagtccacaaataacaaactgtcc 25

Scoring table: IDENTITY.NUC
               Gapop 10.0 , Gapext 1.0

Searched:      2751289 seqs, 1493109276 residues

Total number of hits satisfying chosen parameters:   55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     EST:*
                 1: em_estba:*
                 2: em_esthum:*
                 3: em_estin:*
                 4: em_estmu:*
                 5: em_estov:*
                 6: em_estpl:*
                 7: em_estro:*
                 8: em_htc:*
                 9: em_est1:*
                10: gb_est2:*
                11: gb_htc:*
                12: gb_est3:*
                13: gb_est4:*
                14: gb_est5:*
                15: em_estfun:*
                16: em_estom:*
                17: em_gss_hum:*
                18: em_gss_inv:*
                19: em_gss_pln:*
                20: em_gss_vrt:*
                21: em_gss_fun:*
                22: em_gss_man:*
                23: em_gss_mus:*
                24: em_gss_pro:*
                25: em_gss_rod:*
                26: em_gss_phg:*
                27: em_gss_vrl:*
                28: gb_gss1:*
                29: gb_gss2:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	25	100.0	644	28	BZ249455	BZ249455 CH230-314
	2	19.4	77.6	471	10	BB690229	BB690229 BB690229
	3	19.4	77.6	750	13	CA085229	CA085229 SCEZAM209
c	4	19.4	77.6	804	13	BX075142	BX075142 BX075142
	5	19.4	77.6	1109	28	CC277744	CC277744 CH261-95B
	6	19.2	76.8	490	28	AZ027906	AZ027906 RPCI-23-3
	7	19.2	76.8	502	28	AQ838510	AQ838510 HS_5011_A
c	8	19.2	76.8	640	29	CE568205	CE568205 tigr-gss-
	9	19.2	76.8	707	28	BZ419252	BZ419252 if51c06.b
	10	19.2	76.8	857	28	AZ548099	AZ548099 ENTDE72TF
	11	19.2	76.8	940	28	AZ536025	AZ536025 ENTBC83TF
	12	18.8	75.2	417	14	T75397	T75397 yc89h08.r1
	13	18.8	75.2	476	28	AI501346	AI501346 UI-R-C2p-
	14	18.8	75.2	476	28	B47134	B47134 HS-1068-B1-
	15	18.8	75.2	664	12	BI395174	BI395174 TUDPc160
c	16	18.8	75.2	930	10	BE973161	BE973161 60152046
c	17	18.6	74.4	446	14	CF214662	CF214662 CGF100081
	18	18.6	74.4	456	14	CF214741	CF214741 CGF100082
c	19	18.6	74.4	479	14	CB346046	CB346046 CAB2SG000
	20	18.6	74.4	494	12	BI803368	BI803368 H108D09 E
	21	18.6	74.4	552	28	AQ652543	AQ652543 Sheared D
c	22	18.6	74.4	570	13	BQ985897	BQ985897 QGE8A14.y
c	23	18.6	74.4	594	14	CF114153	CF114153 Shultzomi
	24	18.6	74.4	602	13	BQ795615	BQ795615 EST 4553
c	25	18.6	74.4	609	14	CB916135	CB916135 VVD107A10
	26	18.6	74.4	623	14	CD013354	CD013354 VVC015F11
c	27	18.6	74.4	626	28	AQ953684	AQ953684 Sheared D
c	28	18.6	74.4	636	28	B94941	B94941 CIT-HSP-216
	29	18.6	74.4	641	13	BQ796004	BQ796004 EST 4942
	30	18.6	74.4	663	14	CB981751	CB981751 CAB70005_
	31	18.6	74.4	721	29	CE282804	CE282804 tigr-gss_
	32	18.6	74.4	724	14	CF405984	CF405984 CSECS062C
	33	18.6	74.4	764	28	BH052648	BH052648 RPCI-24-3
	34	18.6	74.4	774	14	CD012425	CD012425 VVC015F11
	35	18.6	74.4	785	13	BQ797479	BQ797479 EST 6417
	36	18.6	74.4	793	14	CF763774	CF763774 CBS005091
	37	18.6	74.4	821	13	BQ797186	BQ797186 EST 6124
c	38	18.6	74.4	826	14	CB981682	CB981682 CAB70005_
	39	18.6	74.4	881	13	BQ797594	BQ797594 EST 6532_
c	40	18.6	74.4	902	13	BU269957	BU269957 60350B418
	41	18.6	74.4	945	29	CG047759	CG047759 PUFKV76TD
	42	18.6	74.4	1173	28	CC195469	CC195469 CH261-72B
c	43	18.4	73.6	173	14	CB931753	CB931753 ri62g05.y
c	44	18.4	73.6	242	9	AA650317	AA650317 ns89g04.s
c	45	18.4	73.6	334	28	BZ652437	BZ652437 OGANT62TC

ALIGNMENTS

```
RESULT 1
BZ249455/c
LOCUS       BZ249455     CH230-314P20.TJ CHORI-230    644 bp    DNA    linear   GSS 12-OCT-2002
DEFINITION  CH230-314P20, genomic survey sequence.
ACCESSION   BZ249455
VERSION     BZ249455.1  GI:23910687
KEYWORDS    GSS.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 644)
AUTHORS     Zhao,S., Shetty,J., Shatsman,S., Gebregeorgis,E., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
TITLE       Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL     Unpublished (1999)
COMMENT     Other_GSSs: CH230-314P20.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
```

Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.htm). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or ering information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 314 row: P column: 20
Seq primer: SP6

FEATURES
  source   Class: BAC ends.
                  Location/Qualifiers
                  1..644
                  /organism="Rattus norvegicus"
                  /mol_type="genomic DNA"
                  /strain="BN/SsNHsd/MCW"
                  /db_xref="taxon:10116"
                  /clone="CH230-314P20"
                  /sex="Female"
                  /cell_type="Brain"
                  /clone_lib="CHORI-230 Segment 2"
                  /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
                  CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                  Pieter de Jong"

ORIGIN

Query Match   100.0%;   Score 25; DB 28;   Length 644;
Best Local Similarity 100.0%;   Pred. No. 97;
Matches   25; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 CAGTCCACAAATAAACAAACTGTCC 25
           |||||||||||||||||||||||||
Db   162 CAGTCCACAAATAAACAAACTGTCC 138

RESULT 2
BB690229
BB690229   471 bp   mRNA   linear   EST 10-OCT-2001
LOCUS   BB690229 RIKEN full-length enriched, 12 days embryo female
DEFINITION mullerian duct Mus musculus cDNA clone 6820446O11 3', mRNA
           sequence.
ACCESSION BB690229
VERSION   BB690229.1 GI:16016962
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 471)
AUTHORS   Akimura,T., Arakawa,T., Carninci,P., Furuno,M., Hanagaki,T.,
           Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T., Imotani,K.,
           Ishii,Y., Ito,M., Kawai,J., Kojima,Y., Konno,H., Kouda,M.,
           Matsuyama,T., Nakamura,M., Nishi,K., Nomura,K., Numasaki,R.,
           Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sakazume,N.,
           Sasaki,D., Sato,K., Shibata,K., Shinagawa,A., Shiraki,T.,
           Sogabe,Y., Suzuki,H., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
           Tanaka,T., Tomaru,A., Toya,T., Watahiki,A., Yasunishi,A.,
           Muramatsu,M. and Hayashizaki,Y.
TITLE   RIKEN Encyclopedia of Mouse Full-length cDNAs (Akimura,T., et al.
           2001)
JOURNAL   Unpublished (2001)
COMMENT   Contact: Yoshihide Hayashizaki
           Laboratory for Genome Exploration Research Group, RIKEN Genomic
           Sciences Center (GSC), Yokohama Institute
           The Institute of Physical and Chemical Research (RIKEN)
           1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
           Tel: 81-45-503-9222
           Fax: 81-45-503-9216
           Email: genome-res@gsc.riken.go.jp,
           URL:http://genome.gsc.riken.go.jp/,
           Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
           Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
           Normalization and subtraction of cap-trapper-selected cDNAs to
           prepare full-length cDNA libraries for rapid discovery of new
           genes. Genome Res. .10 (10), 1617-1630 (2000)

wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T.,
Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M. Inoue,Y., Kira,A.
and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P.,
Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. .11 (2), 281-289 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for
further details.
e mouse tissues.

FEATURES
  source   Location/Qualifiers
                  1..471
                  /organism="Mus musculus"
                  /mol_type="mRNA"
                  /strain="C57BL/6J"
                  /db_xref="taxon:10090"
                  /clone="6820446O11"
                  /sex="female"
                  /tissue_type="mullerian duct includes surrounding region"
                  /dev_stage="12 days embryo"
                  /lab_host="DH10B"
                  /clone_lib="RIKEN full-length enriched, 12 days embryo
                  female mullerian duct"
                  /note="Site 1: SalI; Site 2: BamHI; cDNA library was
                  prepared and sequenced in Mouse Genome Encyclopedia
                  Project of Genome Exploration Research Group in Riken
                  Genomic Sciences Center and Genome Science Laboratory in
                  RIKEN. Division of Experimental Animal Research in Riken
                  contributed to prepare mouse tissues. 1st strand cDNA was
                  primed with a primer (5'
                  GAGAGAGAAGAGATCCAAGAGCTCTTTTTTTTTTTTTTTTVN 3'], cDNA was
                  prepared by using trehalose thermo-activated reverse
                  transcriptase and subsequently enriched for full-length by
                  cap-trapper. Second strand cDNA was prepared with the
                  primer adapter of sequence [5'
                  GAGAGAGAGATTCTCGAGTTAATTAAATTAARTCCCCCCCCCCCC 3']. cDNA
                  was cloned into the XhoI and BamHI sites. Vector: a
                  modified pBluescript KS(+) after bulk excision from Lambda
                  FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI"

ORIGIN

Query Match   77.6%;   Score 19.4; DB 10;   Length 471;
Best Local Similarity 95.2%;   Pred. No. 7.3e-03;
Matches   20; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 CAGTCCACAAATAAACAAACT 21
           |||||||||||||||||||||
Db   155 CACTCCACAAATAAACAAACT 175

RESULT 3
CA085229
CA085229
LOCUS   SCEZAM2096B03.g AM2 Saccharum officinarum cDNA clone SCEZAM2096B03
DEFINITION 5', mRNA sequence.
ACCESSION CA085229
VERSION   CA085229.1 GI:34938536
KEYWORDS   EST.
SOURCE   Saccharum officinarum
ORGANISM   Saccharum officinarum
           Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
           Spermatophyta; Magnoliophyta; liliopsida; Poales; Poaceae; PACCAD
           clade; Panicoideae; Andropogoneae; Saccharum.
REFERENCE 1 (bases 1 to 750)
AUTHORS   Vettore,A.L., da Silva,F.R., Kemper,E.L. and Arruda,P.
TITLE   The libraries that made SUCEST
JOURNAL   Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT   Contact: Arruda P

CA085229
SCEZAM2096B03.g AM2 Saccharum officinarum cDNA clone SCEZAM2096B03
5', mRNA sequence.
CA085229
CA085229.1 GI:34938536
EST.
Saccharum officinarum
Saccharum officinarum

750 bp   mRNA   linear   EST 23-SEP-2003

Centro de Biologia Molecular e Engenharia Genetica
Universidade Estadual de Campinas
Caixa Postal 6010, 13083-970, Campinas SP, Brazil
Tel: 55 19 3788 1137
Fax: 55 19 3788 1089
Email: parruda@unicamp.br
Clone distribution: clone distribution information can be found through the Brazilian Clone Collection Center (BCCC) at http://www.bcccenter.fcav.unesp.br
Plate: 096 row: B column: 03
Seq primer: T7 Promoter Primer.

```
FEATURES             Location/Qualifiers
     source          1..750
                     /organism="Saccharum officinarum"
                     /mol_type="mRNA"
                     /db_xref="taxon:4547"
                     /clone="SCEZAM2096B03"
                     /lab_host="DH10B"
                     /clone_lib="AM2"
                     /note="Organ: Apical meristem and tissues surrounding of
                     immature plants; Vector: pSport1; Site_1: SalI; Site_2:
                     NotI; An unidirectional cDNA library generated from
                     [Apical meristem and tissues surrounding of immature
                     plants]. cDNA was prepared from polyA+ mRNA using
                     SuperScript Plasmid System Kit (Invitrogen). The
                     double-strand cDNAs were fractionated in a sepharose
                     CL-2B 40cm-columns and fragments sizing between 0.8 and
                     1.5 Kb were directionally cloned into the vector. Details
                     of each source of RNA and library construction can be
                     obtained at http://sucest.lad.ic.unicamp.br/public"
ORIGIN
```

Query Match   77.6%;   Score 19.4;   DB 13;   Length 750;
Best Local Similarity 95.2%;   Pred. No. 6.3e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    5 CCACAAATAAACAAACTGTCC 25
        ||||||||||||||||||||
Db  441 CCACAAATAAACAAACTGTTC 461
```

RESULT 4
BX075142/c
```
LOCUS       BX075142      804 bp    mRNA    linear   EST 25-APR-2003
DEFINITION  AGENAE Rainbow trout normalized multi-tissues library
            (tcad) Oncorhynchus mykiss cDNA clone tcad0003a.o.23 5prim, mRNA
            sequence.
ACCESSION   BX075142
VERSION     BX075142.1  GI:27738561
KEYWORDS    EST.
SOURCE      Oncorhynchus mykiss (rainbow trout)
  ORGANISM  Oncorhynchus mykiss
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Actinopterygii; Neopterygii; Teleostei; Euteleostei;
            Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.
REFERENCE   1 (bases 1 to 804)
  AUTHORS   Govoroun,M., Guiguen,Y. and Le Gac,F.
  TITLE     Construction and primary characterization of normalized cDNA
            libraries in rainbow trout, Oncorhynchus mykiss
  JOURNAL   Unpublished (2003)
COMMENT     INRA - SCRIBE
            Contact: Guiguen Y
            Campus de beaulieu, RENNES cedex, 35042, France
            Tel: 02.23.48.50.09
            Fax: 02.23.48.50.20
            Email: Yann.Guiguen@beaulieu.rennes.inra.fr
            Sequence cleaned of vector, adaptator and repetitions. Contact us
            at agenasupport@jouy.inra.fr to obtain the chromatogram of this
            sequence.
            Plate: 0003 row: o column: 23
            Seq primer: T7.
FEATURES             Location/Qualifiers
     source          1..804
                     /organism="Oncorhynchus mykiss"
                     /mol_type="mRNA"
                     /db_xref="taxon:8022"
                     /clone="tcad0003a.o.23"
                     /tissue_type="adipose tissue, blood, brain,
                     differentiating gonads, interrenal, intestine, kidney,
                     liver, muscle, ovary, pituitary, testis"
                     /dev_stage="unknown"
                     /lab_host="DH10B"
                     /clone_lib="AGENAE Rainbow trout normalized multi-tissues
                     library (tcad)"
                     /note="Vector: pT7T3D-pac; Clone distribution : AGENAE
                     Resource centre. Francois PIUMI,
                     Francois.Piumi@jouy.inra.fr, INRA, CEA Radiobiologie et
                     Etude du genome (LREG), Domaine de Vilvert, 78352,
                     Jouy-en-Josas cedex, France"
ORIGIN
```

Query Match   77.6%;   Score 19.4;   DB 13;   Length 804;
Best Local Similarity 95.2%;   Pred. No. 6.2e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy    4 TCCACAAATAAACAAACTGTCC 24
        |||||||||||| |||||||||
Db  382 TCCACAAATAATAAATAACTGTC 362
```

RESULT 5
CC277744
```
LOCUS       CC277744      1109 bp    DNA    linear   GSS 13-MAY-2003
DEFINITION  CH261-95B8_RM1.1 CH261 Gallus gallus genomic clone CH261-95B8,
            genomic survey sequence.
ACCESSION   CC277744
VERSION     CC277744.1  GI:3063245
KEYWORDS    GSS.
SOURCE      Gallus gallus (chicken)
  ORGANISM  Gallus gallus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
            Phasianinae; Gallus.
REFERENCE   1 (bases 1 to 1109)
  AUTHORS   Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
            Warren,W., Graves,T., Mardis,E. and Wilson,R.
  TITLE     Gallus gallus BAC End Reads
  JOURNAL   Unpublished (2003)
COMMENT     Contact: Richard K. Wilson
            Genome Sequencing Center
            Washington University School of Medicine
            Email: submissions@watson.wustl.edu
            Insert Length: 182000   Std Error: 0.00
            Seq primer: RM1 TACGACTCACTATAGGGAGA
            Class: BAC ends
            High quality sequence start: 57
            High quality sequence stop: 552.
FEATURES             Location/Qualifiers
     source          1..1109
                     /organism="Gallus gallus"
                     /mol_type="genomic DNA"
                     /strain="Red Jungle Fowl"
                     /db_xref="taxon:9031"
                     /clone="CH261-95B8"
                     /sex="female"
                     /cell_line="UCD001, inbred 256"
                     /clone_lib="CH261"
                     /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
                     CH261 Female Chicken library - for library and clone
                     ordering information: http://www.chori.org/bacpac"
ORIGIN
```

Query Match   77.6%;   Score 19.4;   DB 28;   Length 1109;
Best Local Similarity 95.2%;   Pred. No. 5.6e+03;
Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
Qy     4 TCCACAAATAAACAAACTGTC 24
         ||||||| ||||||||||||
Db  1043 TCCACAAAAAAACAAACTGTC 1063


RESULT 6
AZ027906
LOCUS       AZ027906    490 bp    DNA    linear   GSS 25-FEB-2000
DEFINITION  RPCI-23-346D20.TJ RPCI-23 Mus musculus genomic clone
            RPCI-23-346D20, genomic survey sequence.
ACCESSION   AZ027906
VERSION     AZ027906.1  GI:7103290
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 490)
AUTHORS     Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S.,
            Akinret,B., Levins,M., Mcgann,S., Tsegaye,G., Geer,K., Krol,M., de
            Jong,P. and Fraser,C.M.
TITLE       Mouse BAC End Sequences from Library RPCI-23
JOURNAL     Unpublished (1999)
COMMENT     Other_GSSs: RPCI-23-346D20.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org
            Clones are derived from the mouse BAC library RPCI-23. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering/frame.htm)
            or from Resea ch Genetics (info@resgen.com). BAC end page:
            http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
            Plate: 346   row: D   column: 20
            Seq primer: SP6
            Class: BAC ends.
FEATURES    Location/Qualifiers
source      1..490
            /organism="Mus musculus"
            /mol_type="genomic DNA"
            /strain="C57BL/6J"
            /db_xref="taxon:10090"
            /clone="RPCI-23-346D20"
            /sex="Female"
            /lab_host="DH10B"
            /clone_lib="RPCI-23"
            /note="Organ: Kidney/Brain; Vector; pBACe3.6; Site_1:
            EcoRI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
            brain genomic DNA was isolated and partially digested
            with a combination of EcoRI and EcoRI Methylase. Size
            selected DNA was cloned into the pBACe3.6 vector at the
            EcoRI sites. The ligation products were transformed into
            DH10B electrocompetent cells (BRL Life Technologies)."
ORIGIN

Query Match          76.8%;   Score 19.2;  DB 28;  Length 490;
Best Local Similarity 87.5%;   Pred. No. 8.4e+03;
Matches  21;  Conservative  0;  Mismatches  3;   Indels  0;   Gaps  0;

Qy     1 CAGTCCACAAATAAACAAACTGTC 24
           ||| ||||||||||||||||||
Db   236 CAGACCACAAATAAACAAATTGAC 259


RESULT 7
AQ838510
LOCUS       AQ838510    502 bp    DNA    linear   GSS 30-AUG-1999
DEFINITION  HS_5011_A2_D06_SP6E RPCI-11 Human Male BAC Library Homo sapiens
            genomic clone Plate=587 Col=12 Row=G, genomic survey sequence.


ACCESSION   AQ838510
VERSION     AQ838510.1  GI:5808384
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 502)
AUTHORS     Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., and
            Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and
            Hood,L.
TITLE       Sequence-tagged connectors: A sequence approach to mapping and
            scanning the human genome
JOURNAL     Proc. Natl. Acad. Sci. U.S.A.  96 (17), 9739-9744 (1999)
MEDLINE     99305689
PUBMED      10449764
COMMENT     Contact: Mahairas GG, Wallace JC, Hood L
            High Throughput Sequencing Center
            University of Washington
            401 Queen Anne Avenue North, Seattle, WA 98109, USA
            Tel: (206) 616-3618
            Fax: (206) 616-3887
            Email: jwallace@u.washington.edu
            Clones are derived from the human BAC library RPCI-11. For BAC
            library availability, please contact Pieter de Jong
            (pieter@dejong.med.buffalo.edu). Clones may be purchased from
            BACPAC Resources (http://bacpac.med.buffalo.edu/ordering_bac.htm)
            or from Resear h Genetics (info@resgen.com). BAC end Web Server:
            http://www.htsc.washington.edu
            Plate: 587   row: G   column: 12
            Seq primer: SP6
            Class: BAC ends
            High quality sequence stop: 502.
FEATURES    Location/Qualifiers
source      1..502
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /clone="Plate=587 Col=12 Row=G"
            /sex="male"
            /clone_lib="RPCI-11 Human Male BAC Library"
            /note="Vector: pBACe3.6; Site_1: EcoRI; Site_2: EcoRI;
            Male blood DNA was isolated from one randomly chosen donor
            and partially digested with a combination of EcoRI and
            EcoRI Methylase. Size selected DNA was cloned into the
            pBACe3.6 vector at EcoRI sites"
ORIGIN

Query Match          76.8%;   Score 19.2;  DB 28;  Length 502;
Best Local Similarity 87.5%;   Pred. No. 8.4e+03;
Matches  21;  Conservative  0;  Mismatches  3;   Indels  0;   Gaps  0;

Qy     2 AGTCCACAAATAAACAAACTGTCC 25
           |||||||||||||||||  |||
Db   375 AGTCAACAAATAAACAAACTATAC 398


RESULT 8
CE568205/c
LOCUS       CE568205    640 bp    DNA    linear   GSS 28-SEP-2003
DEFINITION  tigr-gss-dog-1700032545129 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE568205
VERSION     CE568205.1  GI:36884986
KEYWORDS    GSS.
SOURCE      Canis familiaris (dog)
ORGANISM    Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1 (bases 1 to 640)
AUTHORS     Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
```

TITLE   The dog genome: survey sequencing and comparative analysis
JOURNAL   Science 301 (5641), 1898-1903 (2003)
MEDLINE
PUBMED   22875432
            14512627
COMMENT   Contact: Kirkness EF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES   Location/Qualifiers
  source   1..640
            /organism="Canis familiaris"
            /mol_type="genomic DNA"
            /strain="standard Poodle"
            /db_xref="taxon:9615"
            /clone_lib="Dog Library"
            /note="Site 1: BstXI; Libraries were prepared from
ORIGIN   peripheral blood"

Query Match   76.8%;   Score 19.2; DB 29; Length 640;
Best Local Similarity   87.5%;   Pred. No. 7.7e+03;
Matches   21; Conservative   0; Mismatches   3; Indels   0; Gaps   0;

Qy   1 CAGTCCACAAATAAACAAACTGTC 24
           ||||| ||||||||||||| |||||
Db   24 CAGTCAACAAATAAACAAACATTC 1

RESULT 9
BZ419252
LOCUS   BZ419252   707 bp   DNA   linear   GSS 10-DEC-2002
DEFINITION if51c06.b1 WGS-ZmaysF (DH5a methyl filtered) Zea mays genomic clone
            if51c06 5', genomic survey sequence.
ACCESSION   BZ419252
VERSION   BZ419252.1 GI:26364405
KEYWORDS   GSS.
SOURCE   Zea mays
  ORGANISM Zea mays
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
            clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 707)
  AUTHORS   Rabinowicz,P.D., O'Shaughnessy,A.L., Balija,V., Dedhia,N.,
            Katzenburger,F., King,L., Miller,B., Muller,S., Nascimento,L.,
            Zutavern,T., McCombie,W.R. and Martienssen,R.A.
  TITLE   Genomic shotgun sequences from Zea mays (methyl-filtered)
  JOURNAL   Unpublished (2002)
COMMENT   Contact: W. Richard McCombie
            Lita Annenberg Hazen Genome Sequencing Center
            Cold Spring Harbor Laboratory
            PO Box 100, Cold Spring Harbor, NY 11724, USA
            Tel: 516 367 8884
            Fax: 516 367 8874
            Email: mccombie@cshl.org
            Plate: if51 row: c column: 06
            Seq primer: -21M13UnivFwd
            Class: shotgun
            High quality sequence stop: 707.
FEATURES   Location/Qualifiers
  source   1..707
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /cultivar="B73"
            /db_xref="taxon:4577"
            /clone="if51c06"
            /lab_host="DH5a"
            /clone_lib="WGS-ZmaysF (DH5a methyl filtered)"
            /note="Organ: immature ears; Site 1: Xba I; Site 2: Xba I;
            The vector was digested with XbaI and one nucleotide was

added by fill in in the recessive 3' end. The genomic DNA
was nebulized, end repaired, adaptor ligated and size
fractionated using sephadex. The resulting fragments were
between 0.8 and 3 kb and were cloned into the vector (.x/y
reads in M13mp19, .b/g reads in pUC19). The same ligation
was transformed into DH5a."

ORIGIN

Query Match   76.8%;   Score 19.2; DB 28; Length 707;
Best Local Similarity   87.5%;   Pred. No. 7.5e+03;
Matches   21; Conservative   0; Mismatches   3; Indels   0; Gaps   0;

Qy   1 CAGTCCACAAATAAACAAACTGTC 24
           ||| || |||||||||||||||||
Db   231 CTGTGCAAAATAAACAAACTGTC 254

RESULT 10
AZ548099
LOCUS   AZ548099   857 bp   DNA   linear   GSS 14-NOV-2000
DEFINITION ENTDE72TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
            genomic, genomic survey sequence.
ACCESSION   AZ548099
VERSION   AZ548099.1 GI:11171374
KEYWORDS   GSS.
SOURCE   Entamoeba histolytica
  ORGANISM Entamoeba histolytica
            Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1 (bases 1 to 857)
  AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
  TITLE   Determination of clone end sequences from Entamoeba histolytica
            HM1:IMSS sheared DNA library
  JOURNAL   Unpublished (2000)
COMMENT   Contact: Brendan J Loftus
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0208
            Fax: 301 838 3543
            Email: bjloftus@tigr.org
            Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
            DNA library
            Seq primer: M13-Forward
            Class: shotgun
            High quality sequence start: 93
            High quality sequence stop: 620.
FEATURES   Location/Qualifiers
            1..857
  source   1..857
            /organism="Entamoeba histolytica"
            /mol_type="genomic DNA"
            /strain="HM1:IMSS"
            /db_xref="taxon:5759"
            /clone_lib="Entamoeba histolytica Sheared DNA"
            /note="Vector: pHOS1; Site 1: Bst I; Constructed at The
            Institute for Genomic Research (TIGR), Rockville, MD.
            Genomic DNA isolated from broth cultures of E. histolytica
            using a method described by Clark and Diamond (Clark,
            C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
            method for isolate identification. Exp. Parasitol.
            77:450.). The DNA was mechanically sheared to give a
            tight size distribution (~2 kb). The v + i method used for
            the library construction is described in detail in Smith,
            H.O. and Venter, J.C. (Making small insert libraries for
            whole genome shotgun sequencing projects. In Genome
            Sequencing: A Practical Approach, eds. M. Vaudin and B.
            Barell, Oxford University Press, 1999)."

ORIGIN

Query Match   76.8%;   Score 19.2; DB 28; Length 857;
Best Local Similarity   87.5%;   Pred. No. 7e+03;
Matches   21; Conservative   0; Mismatches   3; Indels   0; Gaps   0;

Qy   2 AGTCCACAAATAAACAAACTGTCC 25

RESULT 11
AZ536025
LOCUS   AZ536025   940 bp   DNA   linear   GSS 03-NOV-2000
DEFINITION   ENTBC83TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
             genomic, genomic survey sequence.
ACCESSION   AZ536025
VERSION   AZ536025.1 GI:11092972
KEYWORDS   GSS.
SOURCE   Entamoeba histolytica
ORGANISM   Entamoeba histolytica
             Eukaryota; Entamoebidae; Entamoeba.
REFERENCE   1 (bases 1 to 940)
AUTHORS   Loftus,B., Van Aken,S. and Fraser,C.
TITLE   Determination of clone end sequences from Entamoeba histolytica
             HM1:IMSS sheared DNA library
JOURNAL   Unpublished (2000)
COMMENT   Contact: Brendan J Loftus
             Department of Eukaryotic Genomics
             The Institute for Genomic Research
             9712 Medical Center Dr., Rockville, MD 20850, USA
             Tel: 301 838 0208
             Fax: 301 838 3543
             Email: bjloftus@tigr.org
             Clones are derived from the Entamoeba histolytica HM1:IMSS sheared
             DNA library
             Seq primer: M13-Forward
             Class: shotgun
             High quality sequence start: 16
             High quality sequence stop: 855.
FEATURES   Location/Qualifiers
     source 1..940
             /organism="Entamoeba histolytica"
             /mol_type="genomic DNA"
             /strain="HM1:IMSS"
             /db_xref="taxon:5759"
             /clone_lib="Entamoeba histolytica Sheared DNA"
             /note="Vector: pHOS1; Site_1: Bst I; Constructed at The
             Institute for Genomic Research (TIGR), Rockville, MD.
             Genomic DNA isolated from broth cultures of E. histolytica
             using a method described by Clark and Diamond (Clark,
             C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
             method for isolate identification. Exp. Parasitol.
             77:450.). The DNA was mechanically sheared to give a
             tight size distribution (~2 kb). The v + i method used for
             the library construction is described in detail in Smith,
             H.O. and Venter, J.C. (Making small insert libraries for
             whole genome shotgun sequencing projects. In Genome
             Sequencing: A Practical Approach, eds. M. Vaudin and B.
             Barell, Oxford University Press, 1999)."

ORIGIN

Query Match   76.8%;   Score 19.2; DB 28;   Length 940;
Best Local Similarity   87.5%;   Pred. No. 6.8e+03;
Matches 21; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy   2 AGTCCACAAATAAACAAACTGTCC 25
            |||||| |||||||||| ||||||
Db   264 AGTCAACAAATAAACAAATTTTCC 287

RESULT 12
T75397
LOCUS   T75397   417 bp   mRNA   linear   EST 03-MAR-1995
DEFINITION   yc89h08.r1 Soares infant brain 1NIB Homo sapiens cDNA clone
             IMAGE:23461 5', mRNA sequence.
ACCESSION   T75397
VERSION   T75397.1 GI:692159
KEYWORDS   EST.
SOURCE   Homo sapiens (human)

ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 417)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
             Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
             Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
             Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
             Wilson,R.
TITLE   The WashU-Merck EST Project
JOURNAL   Unpublished (1995)
COMMENT   Contact: Wilson RK
             Washington University School of Medicine
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: est@watson.wustl.edu
             Insert Size: 1820
             High quality sequence stops: 334 Source: IMAGE Consortium, LLNL
             This clone is available royalty-free through LLNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             Insert Length: 1820   Std Error: 0.00
             Seq primer: M13RP1
             High quality sequence stop: 334.
FEATURES   Location/Qualifiers
     source 1..417
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="GDB:395808"
             /db_xref="taxon:9606"
             /clone="IMAGE:23461"
             /sex="female"
             /dev_stage="73 days post natal"
             /lab_host="DH10B (ampicillin resistant)"
             /clone_lib="Soares infant brain 1NIB"
             /note="Organ: whole brain; Vector: Lafmid BA; Site_1: Not
             I; Site_2: Hind III; 1st strand cDNA was primed with a Not
             I - oligo(dT) primer [5',
             AACTGGAAGAATTCGGCGCCGCAGGAATTTTTTTTTTTTTTTT 3'];
             double-stranded cDNA was ligated to Hind III adaptors
             (Pharmacia), digested with Not I and directionally cloned
             into the Not I and Hind III sites of the Lafmid BA vector.
             Library went through one round of normalization. Library
             constructed by Bento Soares and M.Fatima Bonaldo."

ORIGIN

Query Match   75.2%;   Score 18.8; DB 14;   Length 417;
Best Local Similarity   90.9%;   Pred. No. 1.2e+04;
Matches 20; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy   4 TCCACAAATAAACAAACTGTCC 25
            | |||||| ||||||||||||||
Db   319 TACACAAAAAACAAACTGTCC 340

RESULT 13
AI501346
LOCUS   AI501346   476 bp   mRNA   linear   EST 03-JUL-1999
DEFINITION   UI-R-C2p-ru-f-09-0-UI.s1 UI-R-C2p Rattus norvegicus cDNA clone
             UI-R-C2p-ru-f-09-0-UI 3', mRNA sequence.
ACCESSION   AI501346
VERSION   AI501346.1 GI:4399197
KEYWORDS   EST.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
             Rattus.
REFERENCE   1 (bases 1 to 476)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE   Normalization and subtraction: two approaches to facilitate gene
             discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)

Db   327 AGTCAACAAATAAACAAATTTTCC 350
              |||| |||||||||||| | |||

```
MEDLINE     97044477
PUBMED      8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road , 4156  MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            The sequence contained an oligo-dT track that was present in the
            oligonucleotide that was used to prime the synthesis of first
            strand cDNA and therefore this may represent a bonafide poly A
            tail. The sequence tag present in the cDNA between the NotI site
            and the oligo-dT track served to identify it as a clone from the
            normalized kidney library cDNA Library Preparation: M.B. Soares Lab
            Clone distribution: clones will be available through Research
            Genetics (www.resgen.com) This clone is also available through the
            I.M.A.G.E. Consortium at LLNL (info@image.llnl.gov). IMAGE
            ID=1790280 The following repetitive elements were found in this
            cDNA sequence: 1-24, >AT_rich#Low_complexity 33-171, >LIMA6#LINE/L1
            Seq primer: M13 Forward
            POLYA=No.
FEATURES        Location/Qualifiers
     source     1..476
                /organism="Rattus norvegicus"
                /mol_type="mRNA"
                /strain="Sprague-Dawley"
                /db_xref="taxon:10116"
                /clone="UI-R-C2p-ru-f-09-0-UI"
                /dev_stage="adult"
                /lab_host="DH10B (Life Technologies)"
                /clone_lib="UI-R-C2p"
                /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-C2p
                library is a subtracted library derived from the UI-R-C1
                library, which is a subtracted library derived from the
                UI-R-C0 library. The UI-R-C0 library consisted of a
                mixture of individually tagged normalized libraries
                constructed from rat placenta, adult lung, brain, liver,
                kidney, heart, spleen, ovary, muscle, 8, 12 and 18-day
                embryo. The tag is a string of 3-5 nucleotides present
                between the Not I site and the oligo-dT track which allows
                identification of the library of origin of a clone within
                the mixture.The subtracted library (UI-R-C2p) was
                constructed as follows: PCR amplified cDNA inserts from
                UI-R-C1 clones from which 3' ESTs had been derived was
                used as a driver in a hybridization with the UI-R-C1
                library in the form of single-stranded circles. The
                remaining single-stranded circles (subtracted library) was
                purified by hydroxyapatite column chromatography,
                converted to double-stranded circles and electroporated
                into DH10B bacteria (Life Technologies) to generate the
                UI-R-C2p library. This procedure has been previously
                described (Bonaldo, Lennon and Soares, Genome Research 6:
                791-806, 1996)"
ORIGIN

Query Match         75.2%;  Score 18.8;  DB 9;  Length 476;
Best Local Similarity   90.9%;   Pred. No. 1.2e+04;
Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    2 AGTCCACAAATAAACAAACTGT 23
        |||||||||||||||||| ||||
Db  179 AGTCCACAAATAAACAAATTAT 200

RESULT 14
B47134
LOCUS       B47134                      476 bp    DNA            linear   GSS 21-OCT-1997
DEFINITION  HS-1068-B1-B06-MR.abi CIT Human Genomic Sperm Library C Homo
            sapiens genomic clone Plate=CT 278 Col=11 Row=D, genomic survey
            sequence.
ACCESSION   B47134
```

```
VERSION     B47134.1  GI:2551968
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 476)
  AUTHORS   Mahairas,G.G., Zackrone,K.D., Smith,T., Tipton,S., Schmidt,S.,
            Traicoff,R., Abajian,C., Blanchard,A., West,A. and Hood,L.E.
  TITLE     Construction of a Characterized Clone Resource for Genomic
            Sequencing: Generation and Preliminary Analysis of 20,000 Sequence
            Tagged Connectors
  JOURNAL   Unpublished (1997)
COMMENT     Contact: Mahairas GG, Zackrone KD, Hood L
            University of Washington
            Seattle, WA 98195, USA
            Tel: (206) 616-8744
            Fax: (206) 685-7301
            Email: kzackron@u.washington.edu
            Sequence Tagged Connector
            Plate: CT 278  row: D  column: 11
            Class: BAC ends
            High quality sequence stop: 476.
FEATURES        Location/Qualifiers
     source     1..476
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
                /clone="Plate=CT 278 Col=11 Row=D"
                /sex="M"
                /clone_lib="CIT Human Genomic Sperm Library C"
                /note="Organ: sperm; Vector: pBeloBAC11; BAC Clones in
                E-Coli DH10B"
ORIGIN

Query Match         75.2%;  Score 18.8;  DB 28;  Length 476;
Best Local Similarity   90.9%;   Pred. No. 1.2e+04;
Matches   20;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy    1 CAGTCCACAAATAAACAAACTG 22
        |||||| ||||||||||||||||
Db  112 CAGCCCCAAATAAACAAACTG 133

RESULT 15
BI395174
LOCUS       BI395174                    664 bp    mRNA           linear   EST 17-SEP-2002
DEFINITION  TUDPEc160 5'RACE cDNA Equus caballus cDNA clone TUDPE160, mRNA
            sequence.
ACCESSION   BI395174
VERSION     BI395174.1  GI:1611768
KEYWORDS    EST.
SOURCE      Equus caballus (horse)
  ORGANISM  Equus caballus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
REFERENCE   1  (bases 1 to 664)
  AUTHORS   Pascua,I., Dhar,A.K., Fan,Y., Paradis,M.R., Arruga,M.V. and
            Alcivar-Warren,A.
  TITLE     Isolation of expressed sequence tags from a thoroughbred horse
            (Equus caballus) 5'-RACE cDNA library
  JOURNAL   Anim. Genet. 33 (3), 231-232 (2002)
  MEDLINE   22027061
  PUBMED    12030932
COMMENT     Contact: Alcivar-Warren, A.
            Department of Environmental and Population Health
            Tufts University School of Veterinary Medicine
            200 Westboro Road, North Grafton, MA 01536, USA
            Tel: (508) 839-7970
            Fax: (508) 839-7051
            Email: acacia.warren@tufts.edu
            The cDNA was isolated from blood cells (buffy coat) of a
            Thoroughbred septic foal
```

```
PCR PRimers
FORWARD: 5'-(CUA)4GGCCACGCGTCGACTAGTAC-3'
BACKWARD: 5'-(CAU)4CTGTTTGAGCGGATGAAGGTGAA-3'
Insert Length: 664    Std Error: 0.00
Seq primer: M13 reverse.

FEATURES     Location/Qualifiers
source       1..664
             /organism="Equus caballus"
             /mol_type="mRNA"
             /db_xref="taxon:9796"
             /clone="TUDPE160"
             /tissue_type="Blood cells (buffy coat) from a septic
             female"
             /dev_stage="Foal"
             /lab_host="E. coli DH5"
             /clone_lib="5'RACE cDNA"
             /note="Vector: pAMP1 (Gibco BRL); Obtained by 5'RACE
             following the protocol of Gibco BRL. cDNA was  synthesized
             using a gene-specific (IL1-ra) primer (GSP1:
             5'-CTGTTTGAGCGGATGAAGGT-3') from a genomic sequence. dCTP
             was used to add a homopolymeric tail in the 3'end and the
             tailed cDNA was amplified by PCR using an anchor IL-1ra
             gene-specific primer (GSP2:
             5'-CAUCAUCAUCUGTTTGAGCGGATGAAGGTGAA-3') as the reverse
             and a Universal Anchor Primer (Gibco BRL) as the forward
             primer. cDNA  was cloned through UDG cloning method."

ORIGIN

Query Match     75.2%;    Score 18.8;   DB 12;   Length 664;
Best Local Similarity   90.9%;   Pred. No. 1e+04;
Matches   20;  Conservative   0;  Mismatches   2;  Indels   0;  Gaps   0;

Qy    2 AGTCCACAAATAAACAAACTGT 23
          ||| ||||||||||||||||||
Db  221 AGTGAACAAATAAACAAACTGT 242

Search completed: March 25, 2004, 15:30:13
Job time : 192.926 secs
```

```
                        GenCore version 5.1.6
             Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    March 25, 2004, 10:03:04 ; Search time 18.1912 Seconds
                                       (without alignments)
                                       4297.861 Million cell updates/sec

Title:         US-09-963-285-1_COPY_403_423
Perfect score: 21
Sequence:      1 gggattcctagagggaaggag 21

Scoring table: IDENTITY.NUC
               Gapop 10.0 , Gapext 1.0

Searched:      2458946 seqs, 1861504846 residues

Total number of hits satisfying chosen parameters:    4917892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Published Applications NA:*
        1: /cgn2_6/ptcdata/1/pubpna/US07_PUBCOMB.seq:*
        2: /cgn2_6/ptcdata/1/pubpna/PCT_NEW_PUB.seq:*
        3: /cgn2_6/ptcdata/1/pubpna/US06_NEW_PUB.seq:*
        4: /cgn2_6/ptcdata/1/pubpna/US06_PUBCOMB.seq:*
        5: /cgn2_6/ptcdata/1/pubpna/US07_NEW_PUB.seq:*
        6: /cgn2_6/ptcdata/1/pubpna/PCTUS_PUBCOMB.seq:*
        7: /cgn2_6/ptcdata/1/pubpna/US08_NEW_PUB.seq:*
        8: /cgn2_6/ptcdata/1/pubpna/US08_PUBCOMB.seq:*
        9: /cgn2_6/ptcdata/1/pubpna/US09_PUBCOMB.seq:*
       10: /cgn2_6/ptcdata/1/pubpna/US09C_PUBCOMB.seq:*
       11: /cgn2_6/ptcdata/1/pubpna/US09_NEW_PUB.seq:*
       12: /cgn2_6/ptcdata/1/pubpna/US10A_PUBCOMB.seq:*
       13: /cgn2_6/ptcdata/1/pubpna/US10B_PUBCOMB.seq:*
       14: /cgn2_6/ptcdata/1/pubpna/US10C_PUBCOMB.seq:*
       15: /cgn2_6/ptcdata/1/pubpna/US10B_PUBCOMB.seq:*
       16: /cgn2_6/ptcdata/1/pubpna/US10_NEW_PUB.seq:*
       17: /cgn2_6/ptcdata/1/pubpna/US60_NEW_PUB.seq:*
       18: /cgn2_6/ptcdata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

C	Result No.	Score	% Match	Query Length	Length	DB	ID	Description
	1	21	100.0	21	6021	9	US-09-963-285-5	Sequence 5, Appli
	2	21	100.0	21	6458	9	US-09-963-285-1	Sequence 1, Appli
c	3	19.4	92.4		575	14	US-10-029-386-2846	Sequence 2846, Ap
c	4	18	85.7		480	15	US-10-027-632-269935	Sequence 269935,
	5	18	85.7		493	15	US-10-027-632-274752	Sequence 274752,
	6	17.8	84.8		9951	16	US-10-257-166-9	Sequence 9, Appli
	7	17.4	82.9		185	12	US-10-424-599-98202	Sequence 98202, A
c	8	17.4	82.9		507	15	US-10-027-632-190023	Sequence 190023,
c	9	17.4	82.9		507	15	US-10-027-632-190024	Sequence 190024,
c	10	17.4	82.9		507	15	US-10-027-632-190025	Sequence 190025,
	11	17.4	82.9		549	9	US-09-734-017A-45	Sequence 45, Appl
	12	16.8	80.0		603	15	US-10-027-632-186	Sequence 186, App
	13	16.8	80.0		603	15	US-10-027-632-187	Sequence 187, App
c	14	16.8	80.0		604	15	US-10-027-632-17929	Sequence 17929, A
	15	16.8	80.0		613	15	US-10-027-632-54479	Sequence 54479, A
	16	16.8	80.0		613	15	US-10-027-632-54480	Sequence 54480, A
	17	16.8	80.0		613	15	US-10-027-632-55335	Sequence 55335, A
	18	16.8	80.0		613	15	US-10-027-632-56157	Sequence 56157, A
	19	16.8	80.0		613	15	US-10-027-632-56158	Sequence 56158, A
	20	16.8	80.0		613	15	US-10-027-632-295612	Sequence 295612,
	21	16.8	80.0		613	15	US-10-027-632-295613	Sequence 295613,
c	22	16.8	80.0		669	15	US-10-027-632-120651	Sequence 120651,
	23	16.8	80.0		766	15	US-10-027-632-55334	Sequence 55334, A
	24	16.8	80.0		817	15	US-10-027-632-10458	Sequence 10458, A
c	25	16.8	80.0		2817	15	US-10-312-273-6	Sequence 6, Appli
c	26	16.8	80.0		32177	10	US-09-764-891-8062	Sequence 8062, Ap
c	27	16.8	80.0		32177	15	US-10-158-034-113	Sequence 113, App
	28	16.8	80.0		1230025	15	US-10-289-762-1	Sequence 1, Appli
c	29	16.4	78.1		355	12	US-10-085-783A-3550	Sequence 3550, Ap
	30	16.4	78.1		355	15	US-10-242-535A-3550	Sequence 3550, Ap
c	31	16.4	78.1		422	10	US-09-918-995-8357	Sequence 8357, Ap
c	32	16.4	78.1		436	14	US-10-060-036-2776	Sequence 2776, Ap
	33	16.4	78.1		618	15	US-10-027-632-104902	Sequence 104902,
	34	16.4	78.1		738	15	US-10-027-632-12804	Sequence 12804, A
c	35	16.4	78.1		1358	9	US-09-822-830A-212	Sequence 212, App
c	36	16.4	78.1		2805	15	US-10-312-273-8	Sequence 8, Appli
c	37	16.4	78.1		78785	10	US-09-978-167-3	Sequence 3, Appli
c	38	16.4	78.1		78953	15	US-10-085-117-31	Sequence 31, Appl
c	39	16.4	78.1		202001	9	US-09-734-674-3	Sequence 3, Appli
c	40	16.4	78.1		202001	14	US-10-274-990-3	Sequence 3, Appli
	41	16.2	77.1		332	9	US-09-563-817-653	Sequence 653, App
	42	16.2	77.1		389	14	US-10-313-669-30	Sequence 30, Appl
c	43	16.2	77.1		479	10	US-09-918-995-31540	Sequence 31540, A
c	44	16.2	77.1		550	12	US-10-085-783A-2496	Sequence 2496, Ap
c	45	16.2	77.1		550	15	US-10-242-535A-2496	Sequence 2496, Ap

ALIGNMENTS

```
RESULT 1
US-09-963-285-5
;Sequence 5, Application US/09963285
; Patent No. US2002090707A1
; GENERAL INFORMATION:
;   APPLICANT: Enerbck, Sven
;   APPLICANT: Krook, Katarina
;   APPLICANT: Rondahl, Lena
;   APPLICANT: Wasserman, Wyeth
;   TITLE OF INVENTION: PROMOTER SEQUENCES
;   FILE REFERENCE: 13425-042001
;   CURRENT APPLICATION NUMBER: US/09/963,285
;   CURRENT FILING DATE: 2001-09-26
;   PRIOR APPLICATION NUMBER: SE 0004102-0
;   PRIOR FILING DATE: 2000-11-09
;   PRIOR APPLICATION NUMBER: US 60/238,897
;   PRIOR FILING DATE: 2000-10-10
;   PRIOR APPLICATION NUMBER: SE 0003435-5
;   PRIOR FILING DATE: 2000-09-26
;   NUMBER OF SEQ ID NOS: 24
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 6021
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1649)...(438)
US-09-963-285-5

Query Match             100.0%;  Score 21;  DB 9;  Length 6021;
Best Local Similarity   100.0%;  Pred. No. 0.75;
Matches    21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGATTCCTAGAGGGAAGGAG 21
          |||||||||||||||||||||
Db    385 GGGATTCCTAGAGGGAAGGAG 405
```

RESULT 2
US-09-963-285-1
; Sequence 1, Application US/09963285
; Patent No. US2002009070A1
; GENERAL INFORMATION:
; APPLICANT: Enerbck, Sven
; APPLICANT: Krook, Katarina
; APPLICANT: Rondahl, Lena
; APPLICANT: Wasserman, Wyeth
; TITLE OF INVENTION: PROMOTER SEQUENCES
; FILE REFERENCE: 13425-042001
; CURRENT APPLICATION NUMBER: US/09/963,285
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: SE 0004102-0
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/238,897
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: SE 0003435-5
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 6458
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (2235)...(3737)
US-09-963-285-1

Query Match   100.0%;   Score 21;   DB 9;   Length 6458;
Best Local Similarity   100.0%;   Pred. No. 0.75;
Matches   21;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   1 GGGATTCCTAGAGGGAAGGAG 21
         |||||||||||||||||||||
Db   403 GGGATTCCTAGAGGGAAGGAG 423

RESULT 3
US-10-029-386-2846/c
; Sequence 2846, Application US/10029386
; Publication No. US2003019704A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR G
; TITLE OF INVENTION: EXPRESSION ANALYSIS TWO
; FILE REFERENCE: AEOMICA-X-2
; CURRENT APPLICATION NUMBER: US/10/029,386
; CURRENT FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 34288
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 2846
;   LENGTH: 575
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   OTHER INFORMATION: MAP TO CHR16.1
;   OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
;   OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
;   OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
;   OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
;   OTHER INFORMATION: SWISSPROT HIT: Q15046, EVALUE 9.00e-52
;   OTHER INFORMATION: NT HIT: D31890.1, EVALUE 1.00e-103
;   OTHER INFORMATION: EST_HUMAN HIT: D58129.1, EVALUE 0.00e+00
US-10-029-386-2846

Query Match   92.4%;   Score 19.4;   DB 14;   Length 575;
Best Local Similarity   95.2%;   Pred. No. 5.1;
Matches   20;   Conservative   0;   Mismatches   1;   Indels   0;   Gaps   0;

Qy   1 GGGATTCCTAGAGGGAAGGAG 21
         |||||||||||||| |||||||
Db   88 GGGATTCCTAGAGGTAAGGAG 68

RESULT 4
US-10-027-632-269935/c
; Sequence 269935, Application US/10027632
; Publication No. US2003020407S A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 269935
;   LENGTH: 480
;   TYPE: DNA
;   ORGANISM: Human
US-10-027-632-269935

Query Match   85.7%;   Score 18;   DB 15;   Length 480;
Best Local Similarity   100.0%;   Pred. No. 26;
Matches   18;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy   4 ATTCCTAGAGGGAAGGAG 21
         ||||||||||||||||||
Db   110 ATTCCTAGAGGGAAGGAG 93

RESULT 5
US-10-027-632-274752
; Sequence 274752, Application US/10027632
; Publication No. US2003020407S A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358

```
;     PRIOR FILING DATE: 1999-09-28
;     PRIOR APPLICATION NUMBER: US 60/146,002
;     PRIOR FILING DATE: 1999-08-09
;     NUMBER OF SEQ ID NOS: 325720
;     SOFTWARE: FastSEQ for Windows Version 4.0
;     SEQ ID NO 274752
;     LENGTH: 493
;     TYPE: DNA
;     ORGANISM: Human
US-10-027-632-274752

Query Match             85.7%;  Score 18;  DB 15;  Length 493;
Best Local Similarity  100.0%;  Pred. No. 26;
Matches   18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        4 ATTCCTAGAGGGAAGGAG 21
            ||||||||||||||||||
Db      371 ATTCCTAGAGGGAAGGAG 388

RESULT 6
US-10-257-166-9
;  Sequence 9, Application US/10257166
;  Publication No. US2004002323OA1
;  GENERAL INFORMATION:
;  APPLICANT: OLEK, Alexander
;  APPLICANT: PIEPENBROCK, Christian
;  APPLICANT: BERLIN, Kurt
;  TITLE OF INVENTION: Method and Nucleic Acids for Analysing the Methylation of
;  TITLE OF INVENTION: Genes Implicated in Pharmacogenomics
;  FILE REFERENCE: 5013.1011
;  CURRENT APPLICATION NUMBER: US/10/257,166
;  CURRENT FILING DATE: 2002-10-07
;  PRIOR APPLICATION NUMBER: PCT/EP01/07470
;     DE 10032529.7
;     DE 10043826.1
;  PRIOR FILING DATE: 2001-06-29
;     2000-06-30
;     2000-09-01
;  NUMBER OF SEQ ID NOS: 178
;  SEQ ID NO 9
;  LENGTH: 9951
;  TYPE: DNA
;  ORGANISM: Artificial Sequence
;  FEATURE:
;  OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-257-166-9

Query Match             84.8%;  Score 17.8;  DB 16;  Length 9951;
Best Local Similarity   90.5%;  Pred. No. 29;
Matches   19;  Conservative   0;  Mismatches   2;  Indels   0;  Gaps   0;

Qy        1 GGGATTCCTAGAGGGAAGGAG 21
            ||||| |||||||||||||||
Db     9036 GGGATTCGGAGAGGGAAGGAG 9056

RESULT 7
US-10-424-599-98202/c
;  Sequence 98202, Application US/10424599
;  Publication No. US2004031072A1
;  GENERAL INFORMATION:
;  APPLICANT: La Rosa Thomas J
;  APPLICANT: Kovalic David K
;  APPLICANT: Zhou Yihua
;  APPLICANT: Cao Yongwei
;  TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;  TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;  FILE REFERENCE: 38-21(53223)B
;  CURRENT APPLICATION NUMBER: US/10/424,599
;  CURRENT FILING DATE: 2003-04-28
;  NUMBER OF SEQ ID NOS: 285684
;  SEQ ID NO 98202


;     LENGTH: 185
;     TYPE: DNA
;     ORGANISM: Glycine max
;     FEATURE:
;     OTHER INFORMATION: Clone ID: PAT_MRT3847_59690C.1
US-10-424-599-98202

Query Match             82.9%;  Score 17.4;  DB 12;  Length 185;
Best Local Similarity   94.7%;  Pred. No. 53;
Matches   18;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy        1 GGGATTCCTAGAGGGAAGG 19
            ||||||| ||||||||||||
Db       78 GGGATTCCAAGAGGGAAGG 60

RESULT 8
US-10-027-632-190023/c
;  Sequence 190023, Application US/10027632
;  Publication No. US2003020407SA9
;  GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION: Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE: 2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
;  PRIOR APPLICATION NUMBER: US 60/193,483
;  PRIOR FILING DATE: 2000-03-29
;  PRIOR APPLICATION NUMBER: US 60/185,218
;  PRIOR FILING DATE: 2000-02-24
;  PRIOR APPLICATION NUMBER: US 60/167,363
;  PRIOR FILING DATE: 1999-11-23
;  PRIOR APPLICATION NUMBER: US 60/156,358
;  PRIOR FILING DATE: 1999-09-28
;  PRIOR APPLICATION NUMBER: US 60/146,002
;  PRIOR FILING DATE: 1999-08-09
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  NUMBER OF SEQ ID NOS: 325720
;  SEQ ID NO 190023
;  LENGTH: 507
;  TYPE: DNA
;  ORGANISM: Human
US-10-027-632-190023

Query Match             82.9%;  Score 17.4;  DB 15;  Length 507;
Best Local Similarity   85.7%;  Pred. No. 51;
Matches   18;  Conservative   1;  Mismatches   2;  Indels   0;  Gaps   0;

Qy        1 GGGATTCCTAGAGGGAAGGAG 21
            |||  ||||||||||||| :
Db      335 GGGTTTCCTAGAGGGAAGCAR 315

RESULT 9
US-10-027-632-190024/c
;  Sequence 190024, Application US/10027632
;  Publication No. US2003020407SA9
;  GENERAL INFORMATION:
;  APPLICANT: Wang, David G.
;  TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;  TITLE OF INVENTION: Polymorphisms in the Human Genome
;  FILE REFERENCE: 108827.129
;  CURRENT APPLICATION NUMBER: US/10/027,632
;  CURRENT FILING DATE: 2002-04-30
;  PRIOR APPLICATION NUMBER: US 60/218,006
;  PRIOR FILING DATE: 2000-07-12
;  PRIOR APPLICATION NUMBER: US 60/198,676
;  PRIOR FILING DATE: 2000-04-20
```

; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 190024
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-190024

Query Match   82.9%; Score 17.4; DB 15; Length 507;
Best Local Similarity   85.7%; Pred. No. 51;
Matches   18; Conservative   1; Mismatches   1; Indels   2; Gaps   0;

Qy   1 GGGATTCCTAGAGGGAAGGAG 21
         |||||||| |||||||| :
Db   335 GGGTTTCCTAGAGGGAAGCAR 315

RESULT 10
US-10-027-632-190025/c
; Sequence 190025, Application US/10027632
; Publication No. US2003020407SA9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 190025
; LENGTH: 507
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-190025

Query Match   82.9%; Score 17.4; DB 15; Length 507;
Best Local Similarity   85.7%; Pred. No. 51;
Matches   18; Conservative   1; Mismatches   1; Indels   2; Gaps   0;

Qy   1 GGGATTCCTAGAGGGAAGGAG 21
         |||||||| |||||||| :
Db   335 GGGTTTCCTAGAGGGAAGCAR 315

RESULT 11
US-09-734-017A-45
; Sequence 45, Application US/09734017A
; Patent No. US2002014222A1

; GENERAL INFORMATION:
; APPLICANT: Lerchl, Jens
; APPLICANT: Renz, Andreas
; APPLICANT: Ehrhardt, Thomas
; APPLICANT: Reindl, Andreas
; APPLICANT: Cirpus, Petra
; APPLICANT: Bischoff, Friedrich
; APPLICANT: Frank, Markus
; APPLICANT: Freund, Annette
; APPLICANT: Duwenig, Elke
; APPLICANT: Schmidt, Ralf-Michael
; TITLE OF INVENTION: Moss genes from Physcomitrella patens encoding proteins involved
; TITLE OF INVENTION: the
; TITLE OF INVENTION: synthesis of amino acids, vitamins, cofactors, nucleotides and
; TITLE OF INVENTION: nucleosides
; FILE REFERENCE: BASF-NAE-1331-99-US
; CURRENT APPLICATION NUMBER: US/09/734,017A
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: 60/171,100
; PRIOR FILING DATE: 1999-12-16
; NUMBER OF SEQ ID NOS: 87
; SOFTWARE: PatentIn Ver. 2.1/WordPerfect
; SEQ ID NO 45
; LENGTH: 549
; TYPE: DNA
; ORGANISM: Physcomitrella patens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2)..(547)
; OTHER INFORMATION: 85_ppprot1_083_g04
US-09-734-017A-45

Query Match   82.9%; Score 17.4; DB 9; Length 549;
Best Local Similarity   94.7%; Pred. No. 51;
Matches   18; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 GGGATTCCTAGAGGGAAGG 19
         |||||||| ||||||||
Db   247 GGGATTCCTGGAGGGAAGG 265

RESULT 12
US-10-027-632-186
; Sequence 186, Application US/10027632
; Publication No. US2003020407SA9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 186
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-186

```
Query Match            80.0%;    Score 16.8;  DB 15;  Length 603;
Best Local Similarity  90.0%;    Pred. No. 1e+02;
Matches   18;  Conservative   0;  Mismatches   2;  Indels   0;  Gaps   0;

Qy     1 GGGATTCCTAGAGGGAAGGA 20
           |||||| |||||||||||
Db   532 GGGATTGCTAGAGGGAAGCA 551
```

RESULT 13
US-10-027-632-187
; Sequence 187, Application US/10027632
; Publication No. US2003020475A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 187
; LENGTH: 603
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-187

```
Query Match            80.0%;    Score 16.8;  DB 15;  Length 603;
Best Local Similarity  90.0%;    Pred. No. 1e+02;
Matches   18;  Conservative   0;  Mismatches   2;  Indels   0;  Gaps   0;

Qy     1 GGGATTCCTAGAGGGAAGGA 20
           |||||| |||||||||||
Db   532 GGGATTGCTAGAGGGAAGCA 551
```

RESULT 14
US-10-027-632-17929/c
; Sequence 17929, Application US/10027632
; Publication No. US2003020475A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17929
; LENGTH: 604
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-17929

```
Query Match            80.0%;    Score 16.8;  DB 15;  Length 604;
Best Local Similarity  90.0%;    Pred. No. 1e+02;
Matches   18;  Conservative   0;  Mismatches   2;  Indels   0;  Gaps   0;

Qy     2 GGATTCCTAGAGGGAAGGAG 21
           ||||| ||||||||||||
Db   302 GGATTCCTAGAGGCTAGGAG 283
```

RESULT 15
US-10-027-632-54479
; Sequence 54479, Application US/10027632
; Publication No. US2003020475A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 54479
; LENGTH: 613
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-54479

```
Query Match            80.0%;    Score 16.8;  DB 15;  Length 613;
Best Local Similarity  90.0%;    Pred. No. 1e+02;
Matches   18;  Conservative   0;  Mismatches   2;  Indels   0;  Gaps   0;

Qy     1 GGGATTCCTAGAGGGAAGGA 20
           |||||| ||||||||||||
Db   537 GGGATTGCTAGAGGGAAGCA 556
```

Search completed: March 25, 2004, 15:52:55
Job time : 19.1912 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 25, 2004, 09:53:09 ; Search time 158.698 Seconds
         (without alignments)
         3951.570 Million cell updates/sec

Title: US-09-963-285-1_COPY_403_423
Perfect score: 21
Sequence: 1 gggatcctagagggaaggag 21

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database : EST:*
            1: em_estba:*
            2: em_esthum:*
            3: em_estin:*
            4: em_estmu:*
            5: em_estcv:*
            6: em_estpl:*
            7: em_estro:*
            8: em_htc:*
            9: em_estl:*
            10: gb_est2:*
            11: gb_htc:*
            12: gb_est3:*
            13: gb_est4:*
            14: gb_est5:*
            15: em_estfun:*
            16: em_estom:*
            17: em_gss_hum:*
            18: em_gss_inv:*
            19: em_gss_pln:*
            20: em_gss_vrt:*
            21: em_gss_fun:*
            22: em_gss_mam:*
            23: em_gss_mus:*
            24: em_gss_pro:*
            25: em_gss_rod:*
            26: em_gss_phg:*
            27: em_gss_vrl:*
            28: gb_gss1:*
            29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	21	100.0	644	28	BZ249455	BZ249455 CH230-314
c 2	19.4	92.4	426	14	D58129	D58129 HUM349D08B
c 3	19.4	92.4	568	12	BM665327	BM665327 UI-E-CK1-
c 4	19.4	92.4	658	13	BU733958	BU733958 UI-E-CK1-

No.	c	ID	DB	Length	%	Score	ID	Description
5	c	CE219158	29	722	92.4	19.4	CE219158	tigr-gss-
6	c	BB520390	10	307	87.6	18.4	BB520390	BB520390
7		AW892967	10	383	87.6	18.4	AW892967	CM3-NN000
8	c	N89865	14	241	84.8	17.8	N89865	zb34h11.s1
9	c	BB163362	10	285	84.8	17.8	BB163362	BB163362
10		BB33442	10	287	84.8	17.8	BB33442	BB33442
11		CA721082	14	447	84.8	17.8	CA721082	wkm2n.pk0
12	c	BJ483108	12	522	84.8	17.8	BJ483108	BJ483108
13		CA593927	14	530	84.8	17.8	CA593927	wpac.pk0
14	c	BQ238208	13	601	84.8	17.8	BQ238208	TaE05007B
15	c	BF350523	10	631	84.8	17.8	BF350523	PM0-HT033
16		BQ875250	13	644	84.8	17.8	BQ875250	QGI708.y
17		BQ245382	13	698	84.8	17.8	BQ245382	TaE1025C
18		BQ841551	13	720	84.8	17.8	BQ841551	WHS4212_B
19		AG035378	29	733	84.8	17.8	AG035378	Pan_trog1
20		AJ613761	9	741	84.8	17.8	AJ613761	AJ613761
21		BF627209	10	819	84.8	17.8	BF627209	HVSMEb000
22		CC227802	28	1260	84.8	17.8	CC227802	CH261-162
23	c	CE582275	29	596	82.9	17.4	CE582275	tigr-gss-
24		BZ108019	28	658	82.9	17.4	BZ108019	CH230-239
25		BX363379	13	1009	82.9	17.4	BX363379	BX363379
26		CD727691	14	597	81.0	17	CD727691	4032569_1
27		CE214329	29	621	81.0	17	CE214329	tigr-gss-
28	c	CA735110	14	197	80.0	16.8	CA735110	wpi1s.pk0
29	c	CG516047	29	216	80.0	16.8	CG516047	OST73785
30	c	CB068446	14	229	80.0	16.8	CB068446	is19g06.x
31		BF789464	10	242	80.0	16.8	BF789464	602103530
32		AV026659	9	245	80.0	16.8	AV026659	AV026659
33	c	AI593772	10	262	80.0	16.8	AI593772	vt72d10.x
34		BF784394	10	271	80.0	16.8	BF784394	602101170
35		BG352243	12	272	80.0	16.8	BG352243	sab97g06.
36		BB067618	12	285	80.0	16.8	BB067618	BB067618
37		AA733226	9	304	80.0	16.8	AA733226	vt72d10.r
38	c	AW236008	10	304	80.0	16.8	AW236008	xn24d06.x
39	c	CE225854	29	398	80.0	16.8	CE225854	tigr-gss-
40		BY444754	13	403	80.0	16.8	BY444754	BY444754
41		BY444902	13	424	80.0	16.8	BY444902	BY444902
42	c	BH142762	28	428	80.0	16.8	BH142762	TDGDS43TH
43		AQ218325	28	430	80.0	16.8	AQ218325	HS_2011_A
44	c	BI963355	12	433	80.0	16.8	BI963355	id27g06.x
45	c	AZ588700	28	436	80.0	16.8	AZ588700	1M0397N11

ALIGNMENTS

RESULT 1
BZ249455/c
LOCUS   BZ249455   644 bp   DNA   linear   GSS 12-OCT-2002
DEFINITION CH230-314P20.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-314P20, genomic survey sequence.
ACCESSION   BZ249455
VERSION   BZ249455.1 GI:2910687
KEYWORDS   GSS.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 644)
AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
TITLE   Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL   Unpublished (1999)
COMMENT   Other GSSs: CH230-314P20.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org

Clones are derived from the rat BAC library CHORI-230
(http://www.chori.org/bacpac/rat230.html). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/ordering information.htm). BAC end
page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
Plate: 314 row: P column: 20
Seq primer: SP6
Class: BAC ends.

```
FEATURES
     source          Location/Qualifiers
                     1..644
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-314P20"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"

ORIGIN
```

Query Match   100.0%;   Score 21;   DB 28;   Length 644;
Best Local Similarity   100.0%;   Pred. No. 53;
Matches 21;   Conservative 0;   Mismatches 0;   Indels 0;   Gaps 0;

```
Qy     1 GGGATTCCTAGAGGGAAGGAG 21
         |||||||||||||||||||||
Db   135 GGGATTCCTAGAGGGAAGGAG 115
```

RESULT 2
D58129/c
LOCUS   D58129   426 bp   mRNA   linear   EST 28-AUG-1995
DEFINITION HUM349D08B Clontech human aorta polyA+ mRNA (#6572) Homo sapiens
            cDNA clone GEN-349D08 5', mRNA sequence.
ACCESSION   D58129
VERSION   D58129.1 GI:964751
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 426)
AUTHORS   Fujiwara,T., Hirano,H., Katagiri,T., Kawai,A., Kuga,Y., Nagata,M.,
            Okuno,S., Ozaki,K., Shimizu,F., Shimada,Y., Shinomiya,H.,
            Takaichi,A., Takeda,S., Watanabe,T., Takahashi,E., Hirai,Y.,
            Maekawa,H., Shin,S. and Nakamura,Y.
TITLE   Fujiwara et al. (1995)
JOURNAL   Unpublished (1995)
COMMENT   Contact: Tsutomu Fujiwara
            Otsuka GEN Research Institute
            Otsuka Pharmaceutical Co.,Ltd
            463-10 kagasuno Kawauchi-cho, Tokushima, Tokushima, 771-01 Japan
            Tel: 0886-65-2888
            Fax: 0886-37-1035.

```
FEATURES
     source          Location/Qualifiers
                     1..426
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="GEN-349D08"
                     /clone_lib="Clontech human aorta polyA+ mRNA (#6572)"

ORIGIN
```

Query Match   92.4%;   Score 19.4;   DB 14;   Length 426;
Best Local Similarity   95.2%;   Pred. No. 2.6e+02;
Matches 20;   Conservative 0;   Mismatches 1;   Indels 0;   Gaps 0;

```
Qy     1 GGGATTCCTAGAGGGAAGGAG 21
         ||||||||||||| |||||||
Db    68 GGGATTCCTAGAGGTAAGGAG 48
```

RESULT 3
BM665327/c
LOCUS   UI-E-CK1-afl-b-15-0-UI.s1 UI-E-CK1 Homo sapiens 568 bp mRNA linear EST 27-FEB-2002
DEFINITION UI-E-CK1-afl-b-15-0-UI.s1 UI-E-CK1 Homo sapiens cDNA clone
            UI-E-CK1-afl-b-15-0-UI 3', mRNA sequence.
ACCESSION   BM665327
VERSION   BM665327.1 GI:18972136
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 568)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE   Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   97044477
PUBMED   8889548
COMMENT   Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
            cDNA library preparation: Dr. M. Bento Soares, Univeristy of Iowa
            cDNA library Arrayed by: Dr. M. Bento Soares, Univeristy of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, Univeristy of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            The following repetitive elements were found in this cDNA
            sequence: 1-51, >POLY A#Simple_repeat (matched compliment)
            Seq primer: M13 Forward
            POLYA=Yes.

FEATURES
     source   Location/Qualifiers
                     1..568
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-E-CK1-afl-b-15-0-UI"
                     /tissue_type="Retina Foveal and Macular"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-E-CK1"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-CK1 is a normalized cDNA library containing the
                     following tissue(s): Retina Foveal and Macular. The
                     library was constructed according to Bonaldo, Lennon and
                     Soares, Genome Research, 6:791-806, 1996. First strand
                     cDNA synthesis was primed with an oligo-dT primer
                     containing a Not I site. Double stranded cDNA was ligated
                     to an EcoR I adaptor, digested with Not I, and cloned
                     directionally into pT7T3-Pac vector. The oligonucleotide
                     used to prime the synthesis of first-strand cDNA contains
                     a library tag sequence that is located between the Not I
                     site and the (dT)18 tail. The sequence tag for this
                     library is GTCC. This library was created for the program,
                     Gene Discovery in the Visual System, supported by National
                     Eye Institute (NEI).
                     TAG_TISSUE=Foveal and Macular Retina
                     TAG_LIB=UI-E-CK1
                     TAG_SEQ=GTCC"

ORIGIN
```

Query Match     92.4%;   Score 19.4;   DB 12;   Length 568;
Best Local Similarity   95.2%;   Pred. No. 2.7e+02;
Matches  20;   Conservative  0;   Mismatches  1;   Indels  0;   Gaps  0;

```
Qy          1 GGGATTCCTAGAGGGAAGGAG 21
              |||||||||||||| ||||||
Db        383 GGGATTCCTAGAGGTAAGGAG 363
```

Query Match   92.4%; Score 19.4; DB 13; Length 658;
Best Local Similarity 95.2%; Pred. No. 2.8e+02;
Matches   20; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

```
Qy          1 GGGATTCCTAGAGGGAAGGAG 21
              ||||||||||||||| |||||
Db        383 GGGATTCCTAGAGGTAAGGAG 363
```

RESULT 4
BU733958/c
LOCUS   BU733958   658 bp   mRNA   linear   EST 09-OCT-2002
DEFINITION UI-E-CKl-afl-b-15-0-UI.s2 UI-E-CKl Homo sapiens cDNA clone
            UI-E-CKl-afl-b-15-0-UI 3', mRNA sequence.
ACCESSION   BU733958
VERSION   BU733958.1 GI:2361385
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 658)
AUTHORS   Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE   Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL   Genome Res. 6 (9), 791-806 (1996)
MEDLINE   97044477
PUBMED   8889548
COMMENT   Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156   MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            Tissue Procurement: Dr. Gregg Hageman
            cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
            cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
            Clone Distribution: Researchers may obtain clones from Research
            Genetics (www.resgen.com).
            The following repetitive elements were found in this cDNA
            sequence: 1-51, >POLY_A#Simple_repeat (matched compliment)
            Seq primer: M13 FORWARD
            POLYA=Yes.
FEATURES   Location/Qualifiers
     source   1..658
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="UI-E-CKl-afl-b-15-0-UI"
                     /tissue_type="Retina Foveal and Macular"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
                     /clone_lib="UI-E-CKl"
                     /note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a
                     modified polylinker; Site_1: EcoR I; Site_2: Not I;
                     UI-E-CKl is a normalized cDNA library containing the
                     following tissue(s): Retina Foveal and Macular. The
                     library was constructed according to Bonaldo, Lennon and
                     Soares, Genome Research, 6:791-806, 1996. First strand
                     cDNA synthesis was primed with an oligo-dT primer
                     containing a Not I site. Double stranded cDNA was ligated
                     to an EcoR I adaptor, digested with Not I, and cloned
                     directionally into pT7T3-Pac vector. The oligonucleotide
                     used to prime the synthesis of first-strand cDNA contains
                     a library tag sequence that is located between the Not I
                     site and the (dT)18 tail. The sequence tag for this
                     library is GTCC. This library was created for the program,
                     Gene Discovery in the Visual System, supported by National
                     Eye Institute (NEI)."
                     TAG_TISSUB=Foveal and Macular Retina
                     TAG_LIB=UI-E-CK1
                     TAG_SEQ=GTCC"
ORIGIN
```

RESULT 5
CE219158
LOCUS   CE219158   722 bp   DNA   linear   GSS 25-SEP-2003
DEFINITION tigr-gss-dog-1700032687I021 Dog Library Canis familiaris genomic,
            genomic survey sequence.
ACCESSION   CE219158
VERSION   CE219158.1 GI:35374827
KEYWORDS   GSS.
SOURCE   Canis familiaris (dog)
ORGANISM   Canis familiaris
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1 (bases 1 to 722)
AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K.,
            Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and
            Venter,J.C.
TITLE   The dog genome: survey sequencing and comparative analysis
JOURNAL   Science 301 (5641), 1898-1903 (2003)
MEDLINE   22875432
PUBMED   14512627
COMMENT   Contact: Kirkness BF
            The Institute for Genomic Research
            Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
            Rockville, MD 20850, USA
            Tel: 301-838-0200
            Fax: 301-838-0208
            Email: ekirknes@tigr.org
            Class: shotgun.
FEATURES   Location/Qualifiers
     source   1..722
                     /organism="Canis familiaris"
                     /mol_type="genomic DNA"
                     /strain="Standard Poodle"
                     /db_xref="taxon:9615"
                     /clone_lib="Dog Library"
                     /note="Site_1: BstXI; Libraries were prepared from
                     peripheral blood"

```
Qy          1 GGGATTCCTAGAGGGAAGGAG 21
              |||||||||||||||||||||
Db        383 GGGATTCCTAGAGGTAAGGAG 363
```

Query Match   92.4%; Score 19.4; DB 29; Length 722;
Best Local Similarity 95.2%; Pred. No. 2.9e+02;
Matches   20; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

```
Qy          1 GGGATTCCTAGAGGGAAGGAG 21
              |||||||||||||| ||||||
Db        406 GGGATTACTAGAGGAAGGAG 426
```

RESULT 6
BB520390/c
LOCUS   BB520390   307 bp   mRNA   linear   EST 28-JUL-2000
DEFINITION BB520390 RIKEN full-length enriched, 16 days neonate heart Mus
            musculus cDNA clone D830041M21 3', mRNA sequence.
ACCESSION   BB520390
VERSION   BB520390.1 GI:9571848
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 307)
AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T.,
            Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
ORIGIN

Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C.,
Kusakabe,M., Matsuyama,T., Miki,Y., Mizuno,Y., Nakamura,M., Oda,H.,
Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K.,
Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A.,
Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A.,
Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A.,
Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and
Hayashizaki,Y.

TITLE   RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL   Unpublished (2000)
COMMENT   Contact: Yoshihide Hayashizaki
           Laboratory for Genome Exploration Research Group, RIKEN Genomic
           Sciences Center(GSC) Yokohama Institute
           The Institute of Physical and Chemical Research (RIKEN)
           1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
           Tel: 81-45-503-9222
           Fax: 81-45-503-9216
           Email: genome-res@gsc.riken.go.jp;
           URL:http://genome.gsc.riken.go.jp/;
           Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
           Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
           Thermostabilization and thermoactivation of thermolabile enzymes by
           trehalose and its application for the synthesis of full length
           cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
           Itoh,M., Kitsunai,T., Akiyama,J., Shibata,Y., Izawa,M., Kawai,J.,
           Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
           Okazaki,Y. and Hayashizaki,Y.
           Automated filtration-based high-throughput plasmid preparation
           system. Genome Res. 9 (5), 463-470 (1999)
           Carninci,P. and Hayashizaki,Y.
           High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
           19-44 (1999)
           Please visit our web site (http://genome.rtc.riken.go.jp) for
           further details.
FEATURES   Location/Qualifiers
source   1..307
           /organism="Mus musculus"
           /mol_type="mRNA"
           /db_xref="taxon:10090"
           /clone="D83041M21"
           /tissue_type="heart"
           /dev_stage="16 days neonate"
           /lab_host="DH10B"
           /clone_lib="RIKEN full-length enriched, 16 days neonate
           heart"
           /note="Site 1: SalI; Site_2: BamHI; cDNA library was
           prepared and sequenced in Mouse Genome Encyclopedia
           Project of Genome Exploration Research Group in Riken
           Genomic Sciences Center and Genome Science Laboratory in
           RIKEN. Division of Experimental Animal Research in Riken
           contributed to prepare mouse tissues. 1st strand cDNA was
           primed with a primer [5'
           GAGAGAGAGAGAGGGCCCAACTCGAGTTTTTTTTTTTTVN 3'], cDNA was
           prepared by using trehalose thermo-activated reverse
           transcriptase and subsequently enriched for full-length by
           cap-trapper. Second strand cDNA was prepared with the
           primer adapter of sequence [5'
           GAGAGAGAGATCTCGAGTTAATTAAATTAATCCCCCCCCCCCC 3'], cDNA
           was cleaved with BamHI and XhoI. Vector: a modified
           pBluescript KS(+) after bulk excision from Lambda FLC I."
ORIGIN

Query Match   87.6%; Score 18.4; DB 10; Length 307;
Best Local Similarity 95.0%; Pred. No. 6.8e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   2 GGATTCCTAGAGGAAGGAG 21
         ||||||||||||||| |||
Db 262 GGATTCCTAGAGGAAAGAG 243

RESULT 7
AW892967/c
LOCUS   AW892967   383 bp   mRNA   linear   EST 24-MAY-2000
DEFINITION CM3-NN0006-110300-113-c04 NN0006 Homo sapiens cDNA, mRNA sequence.
ACCESSION   AW892967
VERSION   AW892967.1 GI:8057172
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 383)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
TITLE   Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE   20202663
PUBMED   10737800
COMMENT   Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?t1=&t2=CM3-NN0006-110
            300-113-c04&t3=2000-03-11&t4=1)
            Seq primer: puc 18 forward
            High quality sequence start: 11
            High quality sequence stop: 383.
FEATURES   Location/Qualifiers
source   1..383
            /organism="Homo sapiens"
            /mol_type="mRNA"
            /db_xref="taxon:9606"
            /dev_stage="Adult"
            /clone_lib="NN0006"
            /note="Organ: nervous normal; Vector: puc18; Site 1: SmaI;
            Site 2: SmaI; A mini-library was made by cloning products
            derived from ORESTES PCR (U.S. Letters Patent application
            No. 196,716 - Ludwig Institute for Cancer Research)
            profiles into the PUC 18 vector. Reverse transcription of
            tissue mRNA and cDNA amplification were performed under
            low stringency conditions."
ORIGIN

Query Match   87.6%; Score 18.4; DB 10; Length 383;
Best Local Similarity 95.0%; Pred. No. 7.1e+02;
Matches 19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   1 GGGATTCCTAGAGGGAAGGA 20
         | |||||||||||||| |||
Db   46 GAGATTCCTAGAGGGGAAGGA 27

RESULT 8
N89865
LOCUS   N89865   241 bp   mRNA   linear   EST 02-APR-1996
DEFINITION zb34h11.s1 Soares_parathyroid_tumor_NbHPA Homo sapiens cDNA clone
            IMAGE:305541 3', mRNA sequence.
ACCESSION   N89865
VERSION   N89865.1 GI:1443192
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens

REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
              1 (bases 1 to 241)
AUTHORS   Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M.,
              Holman,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M.,
              Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F.,
              Trevaskis,E., Waterston,R., Williamson,A., Wohldmann,P. and
              Wilson,R.
TITLE   The WashU-Merck EST Project
JOURNAL   Unpublished (1995)
COMMENT   Contact: Wilson RK
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              This clone is available royalty-free through LLNL ; contact the
              IMAGE Consortium (info@image.llnl.gov) for further information.
              Seq primer: ET primer
              High quality sequence stop: 183.
FEATURES   Location/Qualifiers
     source   1..241
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="GDB:1248951"
              /db_xref="taxon:9606"
              /clone="IMAGE:305541"
              /tissue_type="parathyroid tumor"
              /dev_stage="adult"
              /lab_host="DH10B (ampicillin resistant)"
              /clone_lib="Soares_parathyroid tumor_NbHPA"
              /note="Organ: parathyroid gland; Vector: pT7T3D
              (Pharmacia) with a modified polylinker; Site 1: Not I;
              Site 2: Eco RI; 1st strand cDNA was primed with a Not I -
              oligo(dT) primer
              [5'-TGTTACCAATTCGAAGTGGGAGCGGCCGCACCAATTTTTTTTTTTTTTTTT
              TTTTT-3'], double-stranded cDNA was size selected, ligated
              to Eco RI adapters (Pharmacia), digested with Not I and
              cloned into the Not I and Eco RI sites of a modified pT7T3
              vector (Pharmacia). Library went through one round of
              normalization to a Cot = 5. Library constructed by Bento
              Soares and M.Fatima Bonaldo. RNA from sporadic parathyroid
              adenomas was kindly provided by Dr. Stephen Marx, National
              Institute of Diabetes and Digestive and Kidney Diseases,
              NIH."

ORIGIN

Query Match   84.8%;   Score 17.8;   DB 14;   Length 241;
Best Local Similarity   90.5%;   Pred. No. 1.2e+03;
Matches   19;   Conservative   0;   Mismatches   2;   Indels   0;   Gaps   0;

Qy   1 GGGATTCCTAGAGGGAAGGAG 21
          |||||||||| |||||||||| |
Db   210 GGGATTCCTTCCTGAGGGAAGGTG 230

RESULT 9
BB163362/c
LOCUS   BB163362   285 bp   mRNA   linear   EST 29-JUN-2000
DEFINITION BB163362 RIKEN full-length enriched, 16 days neonate thymus Mus
            musculus cDNA clone A130072E22 3' similar to AF077858 Mus musculus
            splicing factor SC35 mRNA, mRNA sequence.
ACCESSION   BB163362
VERSION   BB163362.1 GI:8819303
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
  ORGANISM Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 285)
  AUTHORS   Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T.,
            Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
            Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,

            Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
            Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C.,
            Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H.,
            Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K.,
            Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A.,
            Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A.,
            Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A.,
            Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and
            Hayashizaki,Y.
TITLE   RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL   Unpublished (2000)
COMMENT   Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp;
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
            Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Thermostabilization and thermoactivation of thermolabile enzymes by
            trehalose and its application for the synthesis of full length
            cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
            Itoh,M., Kitsunai,T., Akiyama,J., Shibata,K., Izawa,M., Kawai,J.,
            Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
            Okazaki,Y. and Hayashizaki,Y.
            Automated filtration-based high-throughput plasmid preparation
            system. Genome Res. 9 (5), 463-470 (1999)
            Carninci,P. and Hayashizaki,Y.
            High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
            19-44 (1999)
            Please visit our web site (http://genome.rtc.riken.go.jp) for
            further details.
FEATURES   Location/Qualifiers
     source 1..285
             /organism="Mus musculus"
             /mol_type="mRNA"
             /db_xref="taxon:10090"
             /clone="A130072E22"
             /tissue_type="thymus"
             /dev_stage="16 days neonate"
             /lab_host="DH10B"
             /clone_lib="RIKEN full-length enriched, 16 days neonate
             thymus"
             /note="Site_1: SalI; Site_2: BamHI; cDNA library was
             prepared and sequenced in Mouse Genome Encyclopedia
             Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in
             RIKEN. Division of Experimental Animal Research in Riken
             contributed to prepare mouse tissues. 1st strand cDNA was
             primed with a primer [5'
             GAGAGAGAGAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
             prepared by using trehalose thermo-activated reverse
             transcriptase and subsequently enriched for full-length by
             cap-trapper. cDNA went through one round of normalization
             to Rot = 10.0 and subtraction to Rot = 185.0. Second
             strand cDNA was prepared with the primer adapter of
             sequence [5' GAGAGAGAGATTCTCGAGTTAATTAAATTAATCCCCCCCCCCCCCCC
             3']. cDNA was cleaved with XhoI and BamHI. Vector: a
             modified pBluescript KS(+) after bulk excision from Lambda
             FLC I."

ORIGIN

Query Match   84.8%;   Score 17.8;   DB 10;   Length 285;
Best Local Similarity   90.5%;   Pred. No. 1.2e+03;
Matches   19;   Conservative   0;   Mismatches   2;   Indels   0;   Gaps   0;

Qy   1 GGGATTCCTAGAGGGAAGGAG 21
          ||||||||| |||||||||| ||
Db   109 GGCATTCCTAGAGGGAAGGGG 89

```
RESULT 10
BB33442/c
LOCUS           BB33442              287 bp    mRNA    linear    EST 11-JUL-2000
DEFINITION      BB33442 RIKEN full-length enriched, 10 days neonate medulla
                oblongata Mus musculus cDNA clone B830012A04 3', mRNA sequence.
ACCESSION       BB33442
VERSION         BB33442.1  GI:9042205
KEYWORDS        EST.
SOURCE          Mus musculus (house mouse)
ORGANISM        Mus musculus
                Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE       1  (bases 1 to 287)
AUTHORS         Konno,H., Aizawa,K., Akahira,S., Akiyama,J., Arakawa,T.,
                Carninci,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
                Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
                Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
                Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C.,
                Kusakabe,M., Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H.,
                Okazaki,Y., Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K.,
                Shibata,K., Shibata,Y., Shigemoto,Y., Shinagawa,A., Shiraki,T.,
                Sogabe,Y., Sugahara,Y., Suzuki,H., Suzuki,H., Tagawa,A.,
                Takahashi,F., Tominaga,N., Toya,T., Tsunoda,Y., Watahiki,A.,
                Watanabe,S., Yamamura,T., Yamanaka,I., Yano,R., Yasunishi,A.,
                Yokota,T., Yoshida,K., Yoshiki,A., Yoshino,M., Muramatsu,M. and
                Hayashizaki,Y.
TITLE           RIKEN Mouse ESTs (Konno,H., et al.)
JOURNAL         Unpublished (2000)
COMMENT         Contact: Yoshihide Hayashizaki
                Laboratory for Genome Exploration Research Group, RIKEN Genomic
                Sciences Center(GSC), Yokohama Institute
                The Institute of Physical and Chemical Research (RIKEN)
                1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
                Tel: 81-45-503-9222
                Fax: 81-45-503-9216
                Email: genome-res@gsc.riken.go.jp,
                URL:http://genome.gsc.riken.go.jp/
                Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagaoka,S.,
                Sasaki,N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
                Thermostabilization and thermoactivation of thermolabile enzymes by
                trehalose and its application for the synthesis of full length
                cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
                Itoh,M., Kitsunai,T., Akiyama,J., Shibata,M., Izawa,M., Kawai,J.,
                Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M.,
                Okazaki,Y. and Hayashizaki,Y.
                Automated filtration-based high-throughput plasmid preparation
                system. Genome Res. 9 (5), 463-470 (1999)
                Carninci,P. and Hayashizaki,Y.
                High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
                19-44 (1999)
                Please visit our web site (http://genome.rtc.riken.go.jp) for
                further details.
FEATURES             Location/Qualifiers
     source          1..287
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="B830012A04"
                     /tissue_type="medulla oblongata"
                     /dev_stage="10 days neonate"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, 10 days neonate
                     medulla oblongata"
                     /note="Site 1: SalI; Site 2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAAGGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTTVN 3'], cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 20.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGAGATTCTGAGTTAATTAAATTAATCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
ORIGIN
     Query Match     84.8%;  Score 17.8;  DB 10;  Length 287;
     Best Local Similarity   90.5%;  Pred. No. 1.2e+03;
     Matches   19;  Conservative  0;  Mismatches   2;  Indels   0;  Gaps   0;

     Qy      1 GGGATTCCTAGAGGGAAGGAG 21
                |||  |||||||||||||||||
     Db     91 GGGGTTGCTAGAGGGAAGGAG 71

RESULT 11
CA721082
LOCUS           CA721082         447 bp    mRNA         linear    EST 26-NOV-2002
DEFINITION      wkm2n.pk009.m17 wkm2n Triticum aestivum cDNA clone wkm2n.pk009.m17
                5' end, mRNA sequence.
ACCESSION       CA721082
VERSION         CA721082.1  GI:25442875
KEYWORDS        EST.
SOURCE          Triticum aestivum (bread wheat)
ORGANISM        Triticum aestivum
                Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
                Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
                Pooideae; Triticeae; Triticum.
REFERENCE       1  (bases 1 to 447)
AUTHORS         Tingey,S.V., Powell,W., Wolters,P., Dolan,M., Hainey,C., Yuan,Z.,
                Miao,G., Caraher,N. and Hanafey,M.K.
TITLE           DuPont Wheat cDNA Sequence
JOURNAL         Unpublished (2002)
COMMENT         Contact: Scott V. Tingey
                Crop Genetics
                E. I. DuPont de Nemours and Company
                1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
                Tel: 302-631-2602
                Fax: 302-631-2607
                Email: Scott.V.Tingey@USA.dupont.com
                Seq primer: M13.
FEATURES             Location/Qualifiers
     source          1..447
                     /organism="Triticum aestivum"
                     /mol_type="mRNA"
                     /cultivar="Spring wheat"
                     /db_xref="taxon:4565"
                     /clone="wkm2n.pk009.m17"
                     /tissue_type="kernel"
                     /lab_host="DH10B"
                     /clone_lib="wkm2n"
                     /note="Vector: pBluescript SK+; Site 1: EcoRI; Site 2:
                     XhoI; Wheat (Triticum aestivum L.) kernel malted 175 hours
                     at 4 C, normalized"
ORIGIN
     Query Match     84.8%;  Score 17.8;  DB 14;  Length 447;
     Best Local Similarity   90.5%;  Pred. No. 1.4e+03;
     Matches   19;  Conservative  0;  Mismatches   2;  Indels   0;  Gaps   0;

     Qy      1 GGGATTCCTAGAGGGAAGGAG 21
                |||  |||||||||||||||||
     Db    120 GGGATACCTAGCGCGGAAGGAG 140

RESULT 12
BJ483108/c
LOCUS           BJ483108         522 bp    mRNA         linear    EST 23-MAY-2002
```

DEFINITION   BJ483108 K. Sato unpublished cDNA library, strain H602 adult,
             heading stage top three leaves Hordeum vulgare subsp. spontaneum
             cDNA clone bah43b10 3', mRNA sequence.
ACCESSION   BJ483108
VERSION   BJ483108.1 GI:21161562
KEYWORDS   EST.
SOURCE   Hordeum vulgare subsp. spontaneum
  ORGANISM   Hordeum vulgare subsp. spontaneum
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Pooideae; Triticeae; Hordeum.
REFERENCE   1 (bases 1 to 522)
  AUTHORS   Sato,K., Saisho,D. and Takeda,K.
  TITLE   Barley EST sequencing project in NIG and Okayama Univ
  JOURNAL   Unpublished (2002)
COMMENT   Contact: Tadasu Shin-i
             Center For Genetic Resource Information
             National Institute of Genetics
             1111 Yata, Mishima, Shizuoka 411-8540, Japan
             Tel: 81-559-81-6856
             Fax: 81-559-81-6855
             Email: tshini@genes.nig.ac.jp.
FEATURES   Location/Qualifiers
     source 1..522
             /organism="Hordeum vulgare subsp. spontaneum"
             /mol_type="mRNA"
             /strain="H602"
             /sub_species="spontaneum"
             /db_xref="taxon:77009"
             /clone="bah43b10"
             /tissue_type="top three leaves"
             /dev_stage="adult, heading stage"
             /clone_lib="K. Sato unpublished cDNA library, strain H602
             adult, heading stage top three leaves"
ORIGIN

Query Match   84.8%; Score 17.8; DB 12; Length 522;
Best Local Similarity   90.5%; Pred. No. 1.4e+03;
Matches   19; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 GGGATTCCTAGAGGGAAGGAG 21
        |||||||| |||||||
Db 389 GGGATACCTAGCGGGAAGGAG 369

RESULT 13
CA593927
LOCUS   CA593927   530 bp   mRNA   linear   EST 21-NOV-2002
DEFINITION   wpalc.pk004.a21 wpalc Triticum aestivum cDNA clone wpalc.pk004.a21
             5' end, mRNA sequence.
ACCESSION   CA593927
VERSION   CA593927.1 GI:25143946
KEYWORDS   EST.
SOURCE   Triticum aestivum (bread wheat)
  ORGANISM   Triticum aestivum
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Pooideae; Triticeae; Triticum.
REFERENCE   1 (bases 1 to 530)
  AUTHORS   Tingey,S.V., Moore,G., Griffiths,S., Powell,W., Wolters,P.,
             Dolan,M., Hainey,C., Miao,G., Caraher,N. and Hanafey,M.K.
  TITLE   DuPont Wheat cDNA Sequence in collaboration with the John Innes
             Center 1
  JOURNAL   Unpublished (2002)
COMMENT   Contact: Scott V. Tingey
             Crop Genetics
             E. I. DuPont de Nemours and Company
             1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
             Tel: 302-631-2602
             Fax: 302-631-2607
             Email: Scott.V.Tingey@USA.dupont.com
             Seq primer: M13.
FEATURES   Location/Qualifiers

     source 1..530
             /organism="Triticum aestivum"
             /mol_type="mRNA"
             /db_xref="taxon:4565"
             /clone="wpalc.pk004.a21"
             /tissue_type="anthers"
             /lab_host="DH10B"
             /clone_lib="wpalc"
             /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
             XhoI; Wheat (Triticum aestivum) pre-meiotic anthers JIC"
ORIGIN

Query Match   84.8%; Score 17.8; DB 14; Length 530;
Best Local Similarity   90.5%; Pred. No. 1.4e+03;
Matches   19; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 GGGATTCCTAGAGGGAAGGAG 21
        |||||||| |||||||
Db 342 GGGATACCTAGCGGGAAGGAG 362

RESULT 14
BQ238208/c
LOCUS   BQ238208   601 bp   mRNA   linear   EST 03-MAY-2002
DEFINITION   TaE05007B02F TaE05 Triticum aestivum cDNA clone TaE05007B02F, mRNA
             sequence.
ACCESSION   BQ238208
VERSION   BQ238208.1 GI:20434084
KEYWORDS   EST.
SOURCE   Triticum aestivum (bread wheat)
  ORGANISM   Triticum aestivum
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
             Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
             Pooideae; Triticeae; Triticum.
REFERENCE   1 (bases 1 to 601)
  AUTHORS   Cloutier,S.
  TITLE   Wheat functional genomics - Glenlea developing seeds cDNA libraries
  JOURNAL   Unpublished (2002)
COMMENT   Contact: Dr. Sylvie Cloutier
             Cereal Research Centre, Agriculture and Agri-food Canada
             195 Dafoe Rd, Winnipeg, MB, Canada R3T 2M9
             Tel: (204) 983-2340
             Fax: (204) 983-4604
             Email: scloutier@agr.gc.ca
             was cloned directionally, not all sequences generated with reverse
             primer were from the 5' end (same with forward primer and 3' end).
             Average insert size is >2.0 kb
             Plate: 007 row: B column: 02
             Seq primer: M13 Forward.
FEATURES   Location/Qualifiers
     source 1..601
             /organism="Triticum aestivum"
             /mol_type="mRNA"
             /cultivar="Glenlea"
             /db_xref="taxon:4565"
             /clone="TaE05007B02F"
             /tissue_type="developing seeds"
             /dev_stage="5 days after anthesis"
             /lab_host="E. coli DH10B"
             /clone_lib="TaE05"
             /note="Vector: pSPORT-P (Invitrogen Technologies); Site_1:
             NotI; Site_2: MluI; mRNA obtained from wheat seeds of
             cultivar Glenlea 5 days post-anthesis"
ORIGIN

Query Match   84.8%; Score 17.8; DB 13; Length 601;
Best Local Similarity   90.5%; Pred. No. 1.4e+03;
Matches   19; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 GGGATTCCTAGAGGGAAGGAG 21
        |||||||| |||||||
Db 519 GGGATACCTAGCGGGAAGGAG 499

```
RESULT 15
BF350523/c
LOCUS       BF350523                       631 bp     mRNA    linear    EST 22-NOV-2000
DEFINITION  PM0-HT0335-310100-005-c02 HT0335 Homo sapiens cDNA, mRNA sequence.
ACCESSION   BF350523
VERSION     BF350523.1   GI:11309597
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 631)
  AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
            Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
            Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H.,
            Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V.,
            O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
            Simpson,A.J.
  TITLE     Shotgun sequencing of the human transcriptome with ORF expressed
            sequence tags
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
  MEDLINE   20202663
   PUBMED   10737800
COMMENT     Contact: Simpson A.J.G.
            Laboratory of Cancer Genetics
            Ludwig Institute for Cancer Research
            Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
            Brazil
            Tel: +55-11-2704922
            Fax: +55-11-2707001
            Email: asimpson@ludwig.org.br
            This sequence was derived from the FAPESP/LICR Human Cancer Genome
            Project. This entry can be seen in the following URL
            (http://www.ludwig.org.br/scripts/gethtml2.pl?ti=PM0&t2=PM0-HT0335-
            310100-005-c02&t3=2000-01-31&t4=1)
            Seq_primer: puc 18 forward
            High quality sequence start: 16
            High quality sequence stop: 36.
FEATURES             Location/Qualifiers
     source          1..631
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /dev_stage="Adult"
                     /clone_lib="HT0335"
                     /note="Organ: head neck; Vector: puc18; Site_1: SmaI;
                     Site_2: SmaI; A mini-library was made by cloning products
                     derived from ORESTES PCR (U.S. Letters Patent application
                     No. 196,716 - Ludwig Institute for Cancer Research)
                     profiles into the pUC 18 vector. Reverse transcription of
                     tissue mRNA and cDNA amplification were performed under
                     low stringency conditions."
ORIGIN

Query Match            84.8%;  Score 17.8;  DB 10;  Length 631;
Best Local Similarity  90.5%;  Pred. No. 1.5e+03;
Matches    19;  Conservative    0;  Mismatches    2;  Indels    0;  Gaps    0;

Qy       1 GGGATTCCTAGAGGGAAGGAG 21
           || ||||| ||||||||||||
Db     164 GGAATTGCTAGAGGGAAGGAG 144
```

Search completed: March 25, 2004, 15:30:18
Job time : 163.698 secs

OM nucleic - nucleic search, using sw model

```
            GenCore version 5.1.6
       Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:        March 25, 2004, 08:39:03 ; Search time 143.084 Seconds
                                           (without alignments)
                                           6361.316 Million cell updates/sec

Title:          US-09-963-285-1_COPY_403_423
Perfect score:  21
Sequence:       1 gggattcctagagggaaggag 21

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters:       6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :      GenEmbl:*
     1: gb_ba:*
     2: gb_htg:*
     3: gb_in:*
     4: gb_om:*
     5: gb_ov:*
     6: gb_pat:*
     7: gb_ph:*
     8: gb_pl:*
     9: gb_pr:*
    10: gb_ro:*
    11: gb_sts:*
    12: gb_sy:*
    13: gb_un:*
    14: gb_vi:*
    15: em_ba:*
    16: em_fun:*
    17: em_hum:*
    18: em_in:*
    19: em_mu:*
    20: em_om:*
    21: em_or:*
    22: em_ov:*
    23: em_pat:*
    24: em_ph:*
    25: em_pl:*
    26: em_ro:*
    27: em_sts:*
    28: em_un:*
    29: em_vi:*
    30: em_htg_hum:*
    31: em_htg_inv:*
    32: em_htg_other:*
    33: em_htg_mus:*
    34: em_htg_pln:*
    35: em_htg_rod:*
    36: em_htg_mam:*
    37: em_htg_vrt:*
    38: em_sy:*
    39: em_htgo_hum:*
    40: em_htgo_mus:*
    41: em_htgo_other:*
```

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1		21	100.0	6021	10	MMMFHEAD1	Y08222 M.musculus
2	c	21	100.0	168656	9	AC009108	AC009108 Homo sapi
3		21	100.0	178416	10	AC127554	AC127554 Mus muscu
4	c	19.4	92.4	156100	2	AC011934	AC011934 Homo sapi
5		19.4	92.4	192590	2	AC118858	AC118858 Rattus no
6		19.4	92.4	194832	9	AC025287	AC025287 Homo sapi
7		18.4	87.6	110999	9	AC130893	AC130893 Homo sapi
8		18.4	87.6	112153	10	BX000520	BX000520 Mouse DNA
9	c	18.4	87.6	145626	2	AC020721	AC020721 Homo sapi
10		18.4	87.6	173516	2	AP002356	AP002356 Homo sapi
11	c	18.4	87.6	183298	2	AC025076	AC025076 Homo sapi
12		18.4	87.6	212884	9	AP001541	AP001541 Homo sapi
13	c	18.4	87.6	245820	2	AC106477	AC106477 Rattus no
14		18	85.7	169032	9	AC016931	AC016931 Homo sapi
15	c	18	85.7	187214	2	AC137676	AC137676 Mus muscu
16		18	85.7	216498	2	AC123722	AC123722 Mus muscu
17		17.8	84.8	9951	6	AX348314	AX348314 Sequence
18		17.8	84.8	119185	9	BX255916	BX255916 Mouse DNA
19		17.8	84.8	144750	9	AL391499	AL391499 Human DNA
20	c	17.8	84.8	148610	2	AC026175	AC026175 Homo sapi
21	c	17.8	84.8	160184	9	AC015821	AC015821 Homo sapi
22		17.8	84.8	160557	9	AC117945	AC117945 Homo sapi
23	c	17.8	84.8	160713	2	AC007174	AC007174 Homo sapi
24		17.8	84.8	161300	2	AC026169	AC026169 Homo sapi
25		17.8	84.8	161449	9	AC005341	AC005341 Homo sapi
26	c	17.8	84.8	162030	2	AC018601	AC018601 Homo sapi
27		17.8	84.8	165608	2	AC087842	AC087842 Rattus no
28		17.8	84.8	169199	9	AC018812	AC018812 Homo sapi
29		17.8	84.8	170977	9	AP000756	AP000756 Homo sapi
30		17.8	84.8	174896	2	AC022012	AC022012 Homo sapi
31		17.8	84.8	186064	2	AC020752	AC020752 Homo sapi
32	c	17.8	84.8	189097	2	AC022002	AC022002 Homo sapi
33	c	17.8	84.8	193423	9	AC026542	AC026542 Homo sapi
34		17.8	84.8	213197	2	AC026198	AC026198 Homo sapi
35		17.8	84.8	224292	2	AC026873	AC026873 Homo sapi
36		17.8	84.8	230956	9	AC090614	AC090614 Homo sapi
37	c	17.8	84.8	231764	9	AC119470	AC119470 Rattus no
38	c	17.8	84.8	240384	2	AC097041	AC097041 Rattus no
39		17.8	84.8	263668	2	AC126697	AC126697 Rattus no
40	c	17.8	84.8	269916	2	AC132650	AC132650 Rattus no
41		17.4	82.9	565	11	G82157	G82157 S20BP6377FA
42	c	17.4	82.9	84278	2	AC024907_3	Continuation (4 of
43	c	17.4	82.9	87386	9	AC133750	AC133750 Homo sapi
44	c	17.4	82.9	168373	9	AL391683	AL391683 Human DNA
45		17.4	82.9	177314	2	AC131919	AC131919 Mus muscu

ALIGNMENTS

```
RESULT 1
MMMFHEAD1
MMMFHEAD1
LOCUS         MMMFHEAD1          6021 bp    DNA      linear    ROD 14-MAY-1997
DEFINITION    M.musculus MFH-1 gene.
ACCESSION     Y08222
VERSION       Y08222.1  GI:1869968
KEYWORDS      mesenchyme fork head-1 protein; MFH-1 gene.
SOURCE        Mus musculus (house mouse)
  ORGANISM    Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE     1
  AUTHORS     Miura,N., Iida,K., Kakinuma,H., Yang,X.L. and Sugiyama,T.
  TITLE       Isolation of the mouse (MFH-1) and human (FKHL 14) mesenchyme fork
              head-1 genes reveals conservation of their gene and protein
```

Pred. No. is the number of results predicted by chance to have a

```
JOURNAL    structures
           Genomics 41 (3), 489-492 (1997)
MEDLINE    97312712
PUBMED     9161153
REFERENCE  2 (bases 1 to 6021)
AUTHORS    Miura,N.
TITLE      Direct Submission
JOURNAL    Submitted (18-SEP-1996) N. Miura, Akita University School of
           Medicine, Department of Biochemistry, 1-1-1 Hondo, Akita 010, JAPAN
FEATURES            Location/Qualifiers
     source         1..6021
                    /organism="Mus musculus"
                    /mol_type="genomic DNA"
                    /strain="129"
                    /db_xref="taxon:10090"
     mRNA           1486..4345
     gene           2070..3554
                    /gene="MFH-1"
     CDS            2070..3554
                    /gene="MFH-1"
                    /codon_start=1
                    /product="mesenchyme fork head-1 protein"
                    /protein_id="CAA69399.1"
                    /db_xref="GI:1869969"
                    /db_xref="GOA:Q61850"
                    /db_xref="SWISS-PROT:Q61850"
                    /translation="MQARYSVSDPNALGVVPVLSEQNYYRAAGSYGGMASPMCVYSGH
                    PEQYGAGMCRSYAPYHHQPAAPKDLVKPYSYIALITMAIQNAPEKKITLNGIYQPIM
                    DRFPFYRENKQGWQNSIRNLSLNECFVKVPRDDKKPGKGSWTLDPDSYNMFENGSF
                    LRRRRFKKKDVPKDKEERAHLKBPPSTTAKGAPTGTPVADGPKEABKKVVVKSEAAS
                    PALPVITKVETLSPEGALQASPRSASSTPAGSPDGSLPEHHAAAPNGLPGFSVETIMT
                    LRTSPPGGDLSPAAARAGLVVPPLALPYAAAPPAAYTOPCAQGLEAAGSAGYQCSMRA
                    MSLYTGABRPAHVCVPPALDEALSDHPSGPGSPLGALNLAAGOEGALGASGHHHQHHG
                    HLHFQAPPAPQPPPAPQPATQATSWYLNHGGDLSHLPGHTFATQQQTFPNVREMFNS
                    HRLGLDNSSLGESQVSNASCQLPYRATPSLYRHAAPYSYDCTKY"
ORIGIN
     polyA_signal   4320..4325

Query Match         100.0%;    Score 21; DB 10;    Length 6021;
Best Local Similarity  100.0%;    Pred. No. 2.2;
Matches   21;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 GGGATTCCTAGAGGGAAGGAG 21
          |||||||||||||||||||||
Db    385 GGGATTCCTAGAGGGAAGGAG 405

RESULT 2
AC009108/c
LOCUS       AC009108   168656 bp   DNA   linear   PRI 29-OCT-2002
DEFINITION  Homo sapiens chromosome 16 clone RP11-46309, complete sequence.
ACCESSION   AC009108
VERSION     AC009108.10   GI:24418066
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 168656)
AUTHORS     DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
TITLE       Direct Submission
JOURNAL     Direct Submission
REFERENCE   2 (bases 1 to 168656)
AUTHORS     Unpublished
TITLE       DOE Joint Genome Institute.
JOURNAL     Direct Submission
            Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3 (bases 1 to 168656)
AUTHORS     DOE Joint Genome Institute, Stanford Human Genome Center and Los
            Alamos National Laboratory.
TITLE       Direct Submission
JOURNAL     Submitted (29-OCT-2002) DOE Joint Genome Institute, 2800 Mitchell
```

```
COMMENT     Drive, Walnut Creek, CA 94598, USA
            On Oct 29, 2002 this sequence version replaced gi:13786306.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center and Los Alamos
            National Laboratory
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.9% of Sequence;
            Estimated Total Number of Errors is 0.2.
FEATURES            Location/Qualifiers
     source         1..168656
                    /organism="Homo sapiens"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:9606"
                    /chromosome="16"
                    /clone="RP11-46309"
ORIGIN

Query Match         100.0%;    Score 21; DB 9;    Length 168656;
Best Local Similarity  100.0%;    Pred. No. 1.5;
Matches   21;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 GGGATTCCTAGAGGGAAGGAG 21
          |||||||||||||||||||||
Db  78330 GGGATTCCTAGAGGGAAGGAG 78310

RESULT 3
AC127554
LOCUS       AC127554   178416 bp   DNA   linear   ROD 27-NOV-2003
DEFINITION  Mus musculus BAC clone RP24-323K23 from chromosome 8, complete
            sequence.
ACCESSION   AC127554
VERSION     AC127554.4   GI:33457241
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 178416)
AUTHORS     Cordes,M. and Haglund,K.
TITLE       The sequence of Mus musculus BAC clone RP24-323K23
JOURNAL     Unpublished (2001)
REFERENCE   2 (bases 1 to 178416)
AUTHORS     Wilson,R.
TITLE       Sequencing of Mus musculus
JOURNAL     Unpublished (2001)
REFERENCE   3 (bases 1 to 178416)
AUTHORS     McPherson,J.D. and Waterston,R.H.
TITLE       Direct Submission
JOURNAL     Submitted (17-JUL-2002) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   4 (bases 1 to 178416)
AUTHORS     Wilson,R.K.
TITLE       Direct Submission
JOURNAL     Submitted (30-JUL-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   5 (bases 1 to 178416)
AUTHORS     Wilson,R.K.
TITLE       Direct Submission
JOURNAL     Submitted (06-AUG-2003) Genome Sequencing Center, 4444 Forest Park
            Parkway, St. Louis, MO 63108, USA
REFERENCE   6 (bases 1 to 178416)
AUTHORS     Wilson,R.
TITLE       Direct Submission
JOURNAL     Submitted (27-NOV-2003) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Aug 6, 2003 this sequence version replaced gi:33342444.
            -------------- Genome Center
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site: http://genome.wustl.edu
            Contact: submissions@watson.wustl.edu
```

```
---------- Summary Statistics
       Center project name: M_BB0323K23
----------

NOTICE:   This sequence may not represent the entire insert of this
clone.   It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. Wes Warren,
Department of Genetics, Washington University, St. Louis MO. For
additional information about the map position of this sequence, see
http://genome.wustl.edu

SOURCE INFORMATION:
The RPCI-24 BAC Library has been constructed by Pieter de Jong and
coworkers (http://www.chori.org) from male C57BL/6J mouse spleen
and/or brain genomic DNA. The clone and detailed information can be
obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
This sequence is the entire insert of the clone. This clone is
overlapped by AC124170.
```

```
FEATURES          Location/Qualifiers
     source        1..178416
                   /organism="Mus musculus"
                   /mol_type="genomic DNA"
                   /db_xref="taxon:10090"
                   /chromosome="8"
                   /map="8"
                   /clone="RP24-323K23"
                   /clone_lib="RPCI-24"
     repeat_region 1791..1839
                   /rpt_family="Alu"
     repeat_region 3054..3288
                   /rpt_family="B4"
     repeat_region 3834..3845
                   /rpt_family="B4"
     repeat_region 3846..3972
                   /rpt_family="Alu"
     repeat_region 3973..4026
                   /rpt_family="B4"
     misc_feature  6269..8820
                   /note="CpG_island (%GC=69.2, o/e=0.78, #CpGs=269)"
     repeat_region 16848..16980
                   /rpt_family="MIR"
     repeat_region 18385..18494
                   /rpt_family="Alu"
     repeat_region 18974..19317
                   /rpt_family="MaLR"
     repeat_region 19821..20013
                   /rpt_family="B2"
     repeat_region 20188..20538
                   /rpt_family="B4"
     repeat_region 20539..20683
                   /rpt_family="MaLR"
     repeat_region 20684..20799
                   /rpt_family="Alu"
     repeat_region 22901..22957
                   /rpt_family="MaLR"
     repeat_region 22968..23090
                   /rpt_family="ID"
     repeat_region 24384..24441
                   /rpt_family="Alu"
                   /rpt_family="B4"
```

```
     repeat_region 25838..26022
                   /rpt_family="MER1_type"
     repeat_region 27267..27380
                   /rpt_family="Alu"
     repeat_region 28700..28893
                   /rpt_family="B2"
     repeat_region 29950..30049
                   /rpt_family="B4"
     repeat_region 30809..30854
                   /rpt_family="ID"
     repeat_region 30859..30963
                   /rpt_family="Alu"
     repeat_region 30987..31062
                   /rpt_family="ID"
     repeat_region 31157..31294
                   /rpt_family="Alu"
     repeat_region 32192..32321
                   /rpt_family="B4"
     repeat_region 32923..33036
                   /rpt_family="Alu"
     repeat_region 33141..33245
                   /rpt_family="Alu"
     repeat_region 33394..33521
                   /rpt_family="B4"
     repeat_region 33806..33891
                   /rpt_family="B4"
     repeat_region 34643..34976
                   /rpt_family="MaLR"
     repeat_region 35547..35686
                   /rpt_family="B2"
     repeat_region 35701..35839
                   /rpt_family="Alu"
     repeat_region 35873..35982
                   /rpt_family="Alu"
     misc_feature  39796..41776
                   /note="CpG_island (%GC=69.3, o/e=0.84, #CpGs=225)"
     repeat_region 47196..47354
                   /rpt_family="B2"
     repeat_region 47365..47498
                   /rpt_family="Alu"
     repeat_region 47507..47550
                   /rpt_family="B4"
     repeat_region 47706..47840
                   /rpt_family="Alu"
     repeat_region 47716..47892
                   /rpt_family="B4"
     repeat_region 47952..48067
                   /rpt_family="Alu"
     misc_feature  51423..52857
                   /note="CpG_island (%GC=66.5, o/e=0.72, #CpGs=121)"
     repeat_region 53367..53484
                   /rpt_family="Alu"
     repeat_region 54128..54233
                   /rpt_family="Alu"
     repeat_region 55219..55240
                   /rpt_family="ERVK"
     repeat_region 55241..55364
                   /rpt_family="Alu"
     repeat_region 55365..55460
                   /rpt_family="ERVK"
     repeat_region 56094..56379
                   /rpt_family="B4"
     repeat_region 59534..59702
                   /rpt_family="B2"
     repeat_region 60089..60165
                   /rpt_family="ID"
     repeat_region 60513..60709
                   /rpt_family="B2"
     repeat_region 60886..61042
                   /rpt_family="B4"
     repeat_region 61852..61997
                   /rpt_family="MER1_type"
     repeat_region 62247..62549
```

```
        repeat_region    /rpt_family="MaLR"
                         62716..62764
        repeat_region    /rpt_family="ID"
                         65555..65793
        repeat_region    /rpt_family="Alu"
                         65929..66103
        repeat_region    /rpt_family="MIR"
                         67019..67319
        repeat_region    /rpt_family="MER121"
                         67684..67898

Query Match     100.0%;   Score 21;   DB 10;   Length 178416;
Best Local Similarity   100.0%;   Pred. No. 1.5;
Matches    21;   Conservative  0;   Mismatches  0;   Indels  0;   Gaps  0;

Qy      1 GGGATTCCTAGAGGGAAGGAG 21
          |||||||||||||||||||||
Db  38512 GGGATTCCTAGAGGGAAGGAG 38532


RESULT 4
AC011934/c
LOCUS       AC011934       156100 bp    DNA     linear    HTG 12-MAR-2000
DEFINITION  Homo sapiens clone RP11-16C11, WORKING DRAFT SEQUENCE, 15 unordered
            pieces.
ACCESSION   AC011934
VERSION     AC011934.5  GI:7230117
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 156100)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens, clone RP11-16C11
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 156100)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,
            Baldwin,J., Barna,N., Beckerly,R., Boguslavkiy,L., Boukhgalter,B.,
            Brown,A., Castle,A., Colangelo,M., Collins,S., Collymore,A.,
            Cooke,P., DeArellano,K., Dewar,K., Domino,M., Donelan,L., Doyle,M.,
            Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D.,
            Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Lehoczky,J., Lieu,C., Locke,K., Locke,K., Macdonald,P., Marquis,N.,
            McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrim,J.,
            Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            Peterson,K., Pollara,V., Riley,R., Roy,A., Santos,R., Severy,P.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X.,
            Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (16-OCT-1999) Whitehead Institute/ MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On Mar 12, 2000 this sequence version replaced gi:6970386.
            All repeats were identified using RepeatMasker:
            Smit,A.F.A. & Green,P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            -------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            -------- Project Information
            Center project name: L3515
            Center clone name: 16_C_11
            -------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 147739 bases at least Q40
            Consensus quality: 152170 bases at least Q30
            Consensus quality: 153664 bases at least Q20

            Insert size: 163000; agarose-fp
            Insert size: 154700; sum-of-contigs
            Quality coverage: 4.7 in Q20 bases; agarose-fp
            Quality coverage: 4.9 in Q20 bases; sum-of-contigs

            --------------
          * NOTE: This is a 'working draft' sequence. It currently
          * consists of 15 contigs. The true order of the pieces
          * is not known and their order in this sequence record is
          * arbitrary. Gaps between the contigs are represented as
          * runs of N, but the exact sizes of the gaps are unknown.
          * This record will be updated with the finished sequence
          * as soon as it is available and the accession number will
          * be preserved.
          *       1    1060: contig of 1060 bp in length
          *    1061    1160: gap of 100 bp
          *    1161    2774: contig of 1614 bp in length
          *    2775    2874: gap of 100 bp
          *    2875    5568: contig of 2694 bp in length
          *    5569    5668: gap of 100 bp
          *    5669    8685: contig of 3017 bp in length
          *    8686    8785: gap of 100 bp
          *    8786   16002: contig of 7217 bp in length
          *   16003   16102: gap of 100 bp
          *   16103   23355: contig of 7253 bp in length
          *   23356   23455: gap of 100 bp
          *   23456   30787: contig of 7332 bp in length
          *   30788   30887: gap of 100 bp
          *   30888   41139: contig of 10452 bp in length
          *   41140   41439: gap of 100 bp
          *   41440   52179: contig of 10740 bp in length
          *   52180   52279: gap of 100 bp
          *   52280   62766: contig of 10487 bp in length
          *   62767   62866: gap of 100 bp
          *   62867   76787: contig of 13921 bp in length
          *   76788   76887: gap of 100 bp
          *   76888   93898: contig of 17011 bp in length
          *   93899   93998: gap of 100 bp
          *   93999  107072: contig of 13074 bp in length
          *  107073  107172: gap of 100 bp
          *  107173  131104: contig of 23932 bp in length
          *  131105  131204: gap of 100 bp
          *  131205  156100: contig of 24896 bp in length
FEATURES             Location/Qualifiers
     source          1..156100
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone_lib="RPCI-11 Human Male BAC"
                     /clone="RP11-16C11"
     misc_feature    1..1060
                     /note="assembly_fragment"
     misc_feature    1161..2774
                     /note="assembly_fragment"
     misc_feature    2875..5568
                     /note="assembly_fragment"
     misc_feature    5669..8685
                     /note="assembly_fragment"
     misc_feature    8785..16002
                     /note="assembly_fragment"
     misc_feature    16103..23355
                     /note="assembly_fragment"
     misc_feature    23456..30787
                     /note="assembly_fragment"
     misc_feature    30888..41339
                     /note="assembly_fragment"
     misc_feature    41440..52179
                     /note="assembly_fragment"
     misc_feature    52280..62766
                     /note="assembly_fragment"
     misc_feature    62867..76787
                     /note="assembly_fragment"
     misc_feature    76888..93898
                     /note="assembly_fragment"
```

```
misc_feature    93999..107072
                /note="assembly_fragment
                clone_end:T7
                vector_side:right"
misc_feature    107173..131104
                /note="assembly_fragment
                clone_end:SP6
                vector_side:left"
misc_feature    131205..156100
                /note="assembly_fragment"

ORIGIN

Query Match          92.4%;    Score 19.4;    DB 2;    Length 156100;
Best Local Similarity  95.2%;    Pred. No. 11;
Matches   20;    Conservative    0;    Mismatches    1;    Indels    0;    Gaps    0;

Qy       1 GGGATTCCTAGAGGGAAGGAG 21
           ||||||||||||||| ||||||
Db  152079 GGGATTCCTAGAGGTAGGAG 152059


RESULT 5
AC118858
ACCESSION   AC118858
LOCUS       AC118858       192590 bp    DNA     linear    HTG 15-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-248J11, WORKING DRAFT SEQUENCE, 2
            unordered pieces.
ACCESSION   AC118858
VERSION     AC118858.4  GI:25008736
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 192590)
AUTHORS     Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,C.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,A., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Oiarmpunsagoon,A., Pai,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,J., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,N., Ren,Y., Reuter,M., Richards,S., Riggs,S.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,R., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wieczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
TITLE       Direct Submission
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 192590)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (21-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 192590)
AUTHORS     Rat Genome Sequencing Consortium.
TITLE       Direct Submission
JOURNAL     Submitted (15-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 15, 2002 this sequence version replaced gi:22856341.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.

            ---------   Genome Center

            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ---------   Project Information
            Center project name: GVWQ
            Center clone name: CH230-248J11
            ---------   Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 165160 bases at least Q40
            Consensus quality: 166820 bases at least Q30
            Consensus quality: 168087 bases at least Q20
            Estimated insert size: 167652; sum-of-contigs estimation
            Quality coverage: 7x in Q20 bases; sum-of-contigs estimation
            ---------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *   consists of 2 contigs. The true order of the pieces
            *   is not known and their order in this sequence record is
            *   arbitrary. Gaps between the contigs are represented as
            *   runs of N, but the exact sizes of the gaps are unknown.
            *   This record will be updated with the finished sequence
            *   as soon as it is available and the accession number will
            *   be preserved.
            *
            *       1    2625: contig of 2625 bp in length
            *    2626    2725: gap of unknown length
            *    2726  192590: contig of 189865 bp in length.
FEATURES             Location/Qualifiers
     source          1..192590
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-248J11"
```

```
      misc_feature      2726..3958
                         /note="wgs_contig"
      misc_feature      190379..192590
                         /note="wgs_contig"

ORIGIN

      Query Match      92.4%;     Score 19.4;    DB 2;    Length 192590;
      Best Local Similarity   95.2%;    Pred. No. 11;
      Matches  20;   Conservative    0;   Mismatches   1;   Indels   0;   Gaps   0;

      Qy         1 GGGATTCCTAGAGGGAAGGAG 21
                   |||||||||||| |||||||||
      Db     72648 GGGATTCCTAGAGGGGAGGAG 72668

RESULT 6
AC025287
LOCUS       AC025287                 194832 bp   DNA     linear   PRI 26-JAN-2002
DEFINITION  Homo sapiens chromosome 16 clone RP11-490B18, complete sequence.
ACCESSION   AC025287
VERSION     AC025287.8   GI:18376863
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 194832)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 194832)
  AUTHORS   DOE Joint Genome Institute.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-MAR-2000) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3  (bases 1 to 194832)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
REFERENCE   4  (bases 1 to 194832)
  AUTHORS   DOE Joint Genome Institute and Stanford Human Genome Center.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-JAN-2002) DOE Joint Genome Institute, 2800 Mitchell
            Drive, Walnut Creek, CA 94598, USA
COMMENT     On Jan 26, 2002 this sequence version replaced gi:17976465.
            Draft Sequence Produced by DOE Joint Genome Institute
            www.jgi.doe.gov
            Finishing Completed at Stanford Human Genome Center
            www-shgc.stanford.edu
            Quality: Phrap Quality >=40 99.4% of Sequence;
            Estimated Total Number of Errors is 0.6.
FEATURES             Location/Qualifiers
     source          1..194832
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-490B18"

ORIGIN

      Query Match      92.4%;     Score 19.4;    DB 9;    Length 194832;
      Best Local Similarity   95.2%;    Pred. No. 11;
      Matches  20;   Conservative    0;   Mismatches   1;   Indels   0;   Gaps   0;

      Qy         1 GGGATTCCTAGAGGGGAAGGAG 21
                   ||||||||||||||||| ||||
      Db     82178 GGGATTCCTAGAGGTAAGGAG 82198

RESULT 7
AC130893
LOCUS       AC130893                 110999 bp   DNA     linear   PRI 03-SEP-2002
```

```
DEFINITION  Homo sapiens X BAC RP13-926M18 (Roswell Park Cancer Institute Human
            BAC Library) complete sequence.
ACCESSION   AC130893
VERSION     AC130893.3   GI:22657460
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 110999)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
            Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
            Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
            Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
            Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
            Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
            Coyle,M.D., Dathorne,S.R., David,R., Davila,R., Davila,M.L., Davis,C.,
            Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
            Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
            Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
            Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
            Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,J., Garner,T.,
            Garza,N., Gill,R., Gorrel,J.H., Guevara,W., Gunaratne,P., Hale,S.,
            Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,R., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Morgan,M., Morris,S., Moser,M., Neal,D., Newtson,J., Newtson,N.,
            Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
            Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
            Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
            Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
            Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshtari,N.,
            Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
            Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,K., Taylor,K.,
            Tang,H., Tansey,J., Tay,C., Taylor,T., Telfrod,B., Thomas,N.,
            Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
            Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
            Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
            Worley,K., Wu,C., Wu,Y., Wu,Y.P., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 110999)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 110999)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 110999)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Sep 3, 2002 this sequence version replaced gi:22538319.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu
```

```
CLONE LENGTH: This sequence does not necessarily represent the
entire insert of this clone. Overlapping regions of clones are only
sequenced and submitted once, so the sequence for the remainder of
the insert may be found in the record for the adjacent clones.
Overlapping clones are noted at the beginning and end of the
Features listing.

ANNOTATION OF FEATURES:
    STSs are identified using ePCR (Genome Res. 7:541-550) searches
of a local database that includes entries from dbSTS, GDB, and
local mapping efforts.
    Repeats are identified using RepeatMasker (A. Smit and P. Green,
unpublished.) for Human and Mouse sequences.
    Genes and Region of sequence similarity are identified by BLAST
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
EST and cDNA sequences. Genes demonstrate at least two exons
flanked by consensus splice sites that maintained sequence
continuity across the splice junctions. Sequences that are not
identical matches are annotated as similar.

SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
standard of double strand coverage with a minimum of 2 clones and 2
reads with no ambiguities or 2 chemistries with a minimum of 2
clones and 3 reads with no ambiguities. If the sequence quality for
a region does not meet this standard, it will be indicated in the
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
standards - estimated error rate less than 1 per 10,000 bases.
Reports of lowest quality individual bases and measures of base
quality are listed below. Description of the metrics can be found
at URL:
http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
ml.

FEATURES             Location/Qualifiers
     source          1..110999
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /clone="RP13-926M18"
     STS             11719..11937
                     /standard_name="D1S3059"
     STS             14741..14954
                     /standard_name="G44369"
     STS             15930..16107
                     /standard_name="D1S2442"
     misc_feature    99193
                     /note="Sized by PCR and restriction digest - 27 bases
                     missing."
                     /function="force join"
     misc_feature    106291..107062
                     /note="Sized by PCR and restriction digest - 200 bases
                     missing - size of repeat
                     971 bases."
                     /function="unresolved tandem repeat"
     misc_feature    108511..110999
                     /note="overlaps bases 1..2481 of clone AC073583"
                     /function"clone overlap"
     STS             108547..108887
                     /standard_name="D7S3182"

ORIGIN
     Query Match        87.6%;   Score 18.4;   DB 9;  Length 110999;
     Best Local Similarity  95.0%;   Pred. No. 41;
     Matches  19;   Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

     Qy     1 GGGATTCCTAGAGGGAAGGA 20
              ||||||||||||||| |||||
     Db  6684 GGGATTACTAGAGGGAAGGA 6703

RESULT 8
```

```
BX000520
LOCUS       BX000520               112153 bp    DNA     linear   ROD 26-NOV-2003
DEFINITION  Mouse DNA sequence from clone RP23-261N8 on chromosome X, complete
            sequence.
ACCESSION   BX000520
VERSION     BX000520.12  GI:38564122
KEYWORDS    HTG.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 112153)
  AUTHORS   Clark,S.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-NOV-2003) Wellcome Trust Sanger Institute, Hinxton,
            Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
            humquery@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
COMMENT     On Nov 26, 2003 this sequence version replaced gi:32567731.
            Sequence from the Mouse Genome Sequencing Consortium whole genome
            shotgun may have been used to confirm this sequence. Sequence data
            from the whole genome shotgun alone has only been used where it has
            a phred quality of at least 30.
            ------------- Genome Center
            Center: Wellcome Trust Sanger Institute
            Center code: SC
            Web site: http://www.sanger.ac.uk
            Contact: humquery@sanger.ac.uk
            -------------
            During sequence assembly data is compared from overlapping clones.
            Where differences are found these are annotated as variations
            together with a note of the overlapping clone name. Note that the
            variation annotation may not be found in the sequence submission
            corresponding to the overlapping clone, as we submit sequences with
            only a small overlap as described above.
            This sequence was finished as follows unless otherwise noted: all
            regions were either double-stranded or sequenced with an alternate
            chemistry or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by at least
            one plasmid subclone or more than one M13 subclone; and the
            assembly was confirmed by restriction digest, except on the rare
            occasion of the clone being a YAC.
            The following abbreviations are used to associate primary accession
            numbers given in the feature table with their source databases:
            Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
            on the WORMPEP database can be found at
            http://www.sanger.ac.uk/Projects/C_elegans/wormpep RP23-261N8 is
            from the RPCI-23 Mouse BAC Library
            constructed by the group of Pieter de Jong.
            For further details see http://www.chori.org/bacpac/home.htm
VECTOR: pBACe3.6.
FEATURES             Location/Qualifiers
     source          1..112153
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="X"
                     /clone="RP23-261N8"
                     /clone_lib="RPCI-23"

ORIGIN
     Query Match        87.6%;   Score 18.4;   DB 10;  Length 112153;
     Best Local Similarity  95.0%;   Pred. No. 41;
     Matches  19;   Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

     Qy     1 GGGATTCCTAGAGGGAAGGA 20
              |||||||||||||| ||||||
     Db 33551 GGGATTCCTAGAGGGGTAGGA 33570

RESULT 9
AC020721/c
LOCUS       AC020721               145626 bp    DNA     linear   HTG 07-JUL-2000
```

```
DEFINITION   Homo sapiens chromosome 11 clone RP11-485A16, WORKING DRAFT
             SEQUENCE, 11 unordered pieces.
ACCESSION    AC020721
VERSION      AC020721.4  GI:7232163
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 145626)
AUTHORS      Waterston,R.H.
TITLE        The sequence of Homo sapiens clone
JOURNAL      Unpublished
REFERENCE    2  (bases 1 to 145626)
AUTHORS      Waterston,R.H.
TITLE        Direct Submission
JOURNAL      Submitted (08-JAN-2000) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
COMMENT      On Mar 13, 2000 this sequence version replaced gi:7024062.
             ----------- Genome Center -----------
             Center: Washington University Genome Sequencing Center
             Center code: WUGSC
             Web site:http://genome.wustl.edu/gsc/index.shtml
             ---------- Project Information ----------
             Center project name: H_NH0485A16
             ---------- Summary Statistics ----------
             Sequencing vector: M13; 87%
             Sequencing vector: plasmid; 13%
             Chemistry: Dye-primer BT; 87% of reads
             Chemistry: Dye-terminator Big Dye; 13% of reads
             Assembly program: Phrap; version 0.990319
             Consensus quality: 136747 bases at least Q40
             Consensus quality: 139412 bases at least Q30
             Consensus quality: 141193 bases at least Q20
             Insert size: 153000; agarose-fp
             Insert size: 144626; sum-of-contigs
             Quality coverage: 3.75 in Q20 bases; agarose-fp
             Quality coverage: 4.01 in Q20 bases; sum-of-contigs
          *  NOTE: This is a 'working draft' sequence. It currently
          *  consists of 11 contigs. The true order of the pieces
          *  is not known and their order in this sequence record is
          *  arbitrary. Gaps between the contigs are represented as
          *  runs of N, but the exact sizes of the gaps are unknown.
          *  This record will be updated with the finished sequence
          *  as soon as it is available and the accession number will
          *  be preserved.
          *       1: contig of 2629 bp in length
          *    2630: gap of unknown length
          *    2730: contig of 4371 bp in length
          *    7101: gap of unknown length
          *    7200: contig of 6664 bp in length
          *   13864: gap of unknown length
          *   13865: contig of 7675 bp in length
          *   13965: gap of unknown length
          *   21640: contig of 10256 bp in length
          *   21739: gap of unknown length
          *   21740: contig of 14964 bp in length
          *   31995: gap of unknown length
          *   32095: contig of 12964 bp in length
          *   32096: gap of unknown length
          *   47059: gap of unknown length
          *   47060: contig of 12964 bp in length
          *   47159: gap of unknown length
          *   47160: contig of 12964 bp in length
          *   60123: gap of unknown length
          *   60124: contig of 16706 bp in length
          *   60223: gap of unknown length
          *   60224: contig of 16706 bp in length
          *   76929: gap of unknown length
          *   76930: contig of 19709 bp in length
          *   77029: gap of unknown length
          *   77030: contig of 19709 bp in length
          *   96738: gap of unknown length
          *   96739: contig of 24147 bp in length
          *   96838: gap of unknown length
          *   96839: contig of 24147 bp in length
          *  120985: gap of unknown length
          *  120986: contig of 24541 bp in length
          *  121085: gap of unknown length
          *  121086: contig of 24541 bp in length.
FEATURES             Location/Qualifiers
     source          1..145626
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /clone="RP11-485A16"
     misc_feature    1..2629
                     /note="assembly_name:Contig6"
     misc_feature    2730..7100
                     /note="assembly_name:Contig7"
     misc_feature    7201..13864
                     /note="assembly_name:Contig8"
     misc_feature    13965..21639
                     /note="assembly_name:Contig9"
     misc_feature    21740..31995
                     /note="assembly_name:Contig10"
     misc_feature    32096..47059
                     /note="assembly_name:Contig11"
     misc_feature    47160..60123
                     /note="assembly_name:Contig12"
     misc_feature    60224..76929
                     /note="assembly_name:Contig13"
     misc_feature    77030..96738
                     /note="assembly_name:Contig14"
     misc_feature    96839..120985
                     /note="assembly_name:Contig15"
     misc_feature    121086..145626
                     /note="assembly_name:Contig16"
                     /clone_end:SP6
                     /vector_side:right"
ORIGIN

Query Match           87.6%;  Score 18.4;  DB 2;  Length 145626;
Best_Local_Similarity 95.0%;  Pred. No. 40;
Matches 19; Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy        1 GGGATTCCTAGAGGGAAGGA 20
            ||||||||||||||||| |||
Db    16623 GAGATTCCTAGAGGGAAGGA 16604

RESULT 10
AP002356
ACCESSION    AP002356
LOCUS        AP002356              173516 bp      DNA     linear   HTG 31-MAY-2000
DEFINITION   Homo sapiens chromosome 11 clone CTD-3243P17 map 11q, WORKING DRAFT
             SEQUENCE, 20 unordered pieces.
ACCESSION    AP002356
VERSION      AP002356.1  GI:8131620
KEYWORDS     HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 173516)
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
             Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Homo sapiens 173,516 genomic DNA of 11q
JOURNAL      Published Only in DataBase (2000)
REFERENCE    2  (bases 1 to 173516)
AUTHORS      Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
             Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE        Direct Submission
JOURNAL      Submitted (29-MAY-2000) Masahira Hattori, The Institute of Physical
             and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
             Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
             Japan (E-mail:hattori@gsc.riken.go.jp,
             URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
             Fax:81-42-778-9924)
COMMENT      ----------- Genome Center -----------
             Center: RIKEN Genomic Sciences Center(GSC)
             Center code: RIKEN
             Web site: http://hgp.gsc.riken.go.jp/
             Contact: hattori@gsc.riken.go.jp
             ---------- Project Information ----------
             Center project name: HumDraft11
```

```
        Center clone name: CTD-3243P17
        ---------- Summary Statistics
        Sequencing vector: PCR products; 100% of reads
        Chemistry: Dye-terminator ET-amersham; 100% of reads
        Assembly program: Phrap; version 0.990329
        Consensus quality: 162824 bases at least Q40
        Consensus quality: 168195 bases at least Q30
        Consensus quality: 170335 bases at least Q20
        Insert size: 171616; sum-of-contigs
        Quality coverage: 5.04x in Q20 bases; sum-of-contigs
        ----------
        NOTE: This is a 'working draft' sequence. It currently consists of
        20 contigs. The true order of the pieces is not known and their
        order in this sequence record is arbitrary. Gaps between the
        contigs are represented as runs N, but the exact sizes of the gaps
        are unknown. This record will be updated with the finished sequence
        as soon as it is available and the accession number will be
        preserved.

              1  21771  contig of  21771 bp in length
          21872  41497  contig of  19626 bp in length
          41598  57282  contig of  15685 bp in length
          57383  72819  contig of  15437 bp in length
          72920  86720  contig of  13801 bp in length
          86821  97172  contig of  10352 bp in length
          97273 107979  contig of  10707 bp in length
         108080 117581  contig of   9502 bp in length
         117682 125644  contig of   7963 bp in length
         125745 132686  contig of   6942 bp in length
         132787 141286  contig of   8500 bp in length
         141387 147754  contig of   6368 bp in length
         147855 153419  contig of   5565 bp in length
         153520 157078  contig of   3559 bp in length
         157179 162389  contig of   5211 bp in length
         162490 164934  contig of   2445 bp in length
         165035 168246  contig of   3212 bp in length
         168347 170645  contig of   2299 bp in length
         170746 172143  contig of   1398 bp in length
         172244 173516  contig of   1273 bp in length
    * NOTE: This is a 'working draft' sequence. It currently
    * consists of 20 contigs. The true order of the pieces
    * is not known and their order in this sequence record is
    * arbitrary. Gaps between the contigs are represented as
    * runs of N, but the exact sizes of the gaps are unknown.
    * This record will be updated with the finished sequence
    * as soon as it is available and the accession number will
    * be preserved.
    *       1  21771: contig of  21771 bp in length
    *   21772  21871: gap of 100 bp
    *   21872  41497: contig of  19626 bp in length
    *   41498  41597: gap of 100 bp
    *   41598  57282: contig of  15685 bp in length
    *   57283  57382: gap of 100 bp
    *   57383  72819: contig of  15437 bp in length
    *   72820  72919: gap of 100 bp
    *   72920  86720: contig of  13801 bp in length
    *   86721  86820: gap of 100 bp
    *   86821  97172: contig of  10352 bp in length
    *   97173  97272: gap of 100 bp
    *   97273 107979: contig of  10707 bp in length
    *  107980 108079: gap of 100 bp
    *  108080 117581: contig of   9502 bp in length
    *  117582 117681: gap of 100 bp
    *  117682 125644: contig of   7963 bp in length
    *  125645 125744: gap of 100 bp
    *  125745 132686: contig of   6942 bp in length
    *  132687 132786: gap of 100 bp
    *  132787 141286: contig of   8500 bp in length
    *  141287 141386: gap of 100 bp
    *  141387 147754: contig of   6368 bp in length
    *  147755 147854: gap of 100 bp
    *  147855 153419: contig of   5565 bp in length
    *  153420 153519: gap of 100 bp
    *  153520 157078: contig of   3559 bp in length
```

```
    *  157079 157178: gap of 100 bp
    *  157179 162389: contig of   5211 bp in length
    *  162390 162489: gap of 100 bp
    *  162490 164934: contig of   2445 bp in length
    *  164935 165034: gap of 100 bp
    *  165035 168246: contig of   3212 bp in length
    *  168247 168346: gap of 100 bp
    *  168347 170645: contig of   2299 bp in length
    *  170646 170745: gap of 100 bp
    *  170746 172143: contig of   1398 bp in length
    *  172144 172243: gap of 100 bp
    *  172244 173516: contig of   1273 bp in length.

FEATURES             Location/Qualifiers
     source          1..173516
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="CTD-3243P17"
     misc_feature    1..21771
                     /note="assembly_fragment"
     misc_feature    21872..41497
                     /note="assembly_fragment"
     misc_feature    41598..57282
                     /note="assembly_fragment"
     misc_feature    57383..72819
                     /note="assembly_fragment"
     misc_feature    72920..86720
                     /note="assembly_fragment"
     misc_feature    86821..97172
                     /note="assembly_fragment"
     misc_feature    97273..107979
                     /note="assembly_fragment"
     misc_feature    108080..117581
                     /note="assembly_fragment"
     misc_feature    117682..125644
                     /note="assembly_fragment"
     misc_feature    125745..132686
                     /note="assembly_fragment"
     misc_feature    132787..141286
                     /note="assembly_fragment"
     misc_feature    141387..147754
                     /note="assembly_fragment"
     misc_feature    147855..153419
                     /note="assembly_fragment"
     misc_feature    153520..157078
                     /note="assembly_fragment"
     misc_feature    157179..162389
                     /note="assembly_fragment"
     misc_feature    162490..164934
                     /note="assembly_fragment"
     misc_feature    165035..168246
                     /note="assembly_fragment"
     misc_feature    168347..170645
                     /note="assembly_fragment"
     misc_feature    170746..172143
                     /note="assembly_fragment"
     misc_feature    172244..173516
                     /note="assembly_fragment"
ORIGIN

Query Match             87.6%;  Score 18.4;  DB 2;  Length 173516;
Best Local Similarity   95.0%;  Pred. No. 39;
Matches   19;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy          1 GGGATTCCTAGAGGGAAGGA 20
              |||||||||||||||||||
Db     152267 GAGATTCCTAGAGGGGAAGGA 152286

RESULT 11
AC025076/c
```

```
LOCUS       AC025076      183298 bp    DNA     linear   HTG 26-MAY-2000
DEFINITION  Homo sapiens clone RP11-156K16, WORKING DRAFT SEQUENCE, 18
            unordered pieces.
ACCESSION   AC025076
VERSION     AC025076.3  GI:8077051
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 183298)
AUTHORS     Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE       Homo sapiens, clone RP11-156K16
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 183298)
AUTHORS     Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
            Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
            Collymore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
            Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
            Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
            Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
            Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczky,J.,
            Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
            McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
            Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
            Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Olivar,T.M., Oliver,J., Peterson,K., Pierre,N.,
            Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
            Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
            Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
            Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Trigilio,J.,
            Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
            Young,G., Zainoun,J., Zimmer,A. and Zody,M.
TITLE       Direct Submission
JOURNAL     Submitted (04-MAR-2000) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT     On May 25, 2000 this sequence version replaced gi:7272259.
            All repeats were identified using RepeatMasker:
            Smit, A.F.A. & Green, P. (1996-1997)
            http://ftp.genome.washington.edu/RM/RepeatMasker.html
            ---------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
            ---------- Project Information
            Center project name: L7828
            Center clone name: 156_K16
            ---------- Summary Statistics
            Sequencing vector: M13; M77815; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.960731
            Consensus quality: 174315 bases at least Q40
            Consensus quality: 178612 bases at least Q30
            Consensus quality: 180475 bases at least Q20
            Insert size: 182000; agarose-fp
            Insert size: 181598; sum-of-contigs
            Quality coverage: 4.6 in Q20 bases; agarose-fp
            Quality coverage: 4.6 in Q20 bases; sum-of-contigs
            --------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 18 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *      1   1774: contig of 1774 bp in length
            *   1775   1874: gap of 100 bp
            *   1875   4363: contig of 2489 bp in length
            *   4363   4463: gap of 100 bp
            *   4464   7303: contig of 2840 bp in length
            *   7304   7403: gap of 100 bp
            *   7404  10043: contig of 2640 bp in length
            *  10044  10143: gap of 100 bp
            *  10144  14123: contig of 3980 bp in length
            *  14124  14223: gap of 100 bp
            *  14224  18507: contig of 4284 bp in length
            *  18508  18607: gap of 100 bp
            *  18608  21538: contig of 2931 bp in length
            *  21539  21638: gap of 100 bp
            *  21639  27268: contig of 5630 bp in length
            *  27269  27368: gap of 100 bp
            *  27369  33110: contig of 5742 bp in length
            *  33111  33210: gap of 100 bp
            *  33211  39973: contig of 6763 bp in length
            *  39974  40073: gap of 100 bp
            *  40074  45638: contig of 5565 bp in length
            *  45639  45738: gap of 100 bp
            *  45739  55034: contig of 9296 bp in length
            *  55035  55134: gap of 100 bp
            *  55135  62645: contig of 7511 bp in length
            *  62646  62745: gap of 100 bp
            *  62746  77397: contig of 14652 bp in length
            *  77398  77497: gap of 100 bp
            *  77498  93454: contig of 15957 bp in length
            *  93455  93554: gap of 100 bp
            *  93555 110875: contig of 17321 bp in length
            * 110876 110975: gap of 100 bp
            * 110976 128888: contig of 17913 bp in length
            * 128889 128988: gap of 100 bp
            * 128989 183298: contig of 54310 bp in length.
FEATURES             Location/Qualifiers
     source          1..183298
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="RP11-156K16"
                     /clone_lib="RPCI-11 Human Male BAC"
     misc_feature    1..1774
                     /note="assembly_fragment"
     misc_feature    1875..4363
                     /note="assembly_fragment"
     misc_feature    4464..7303
                     /note="assembly_fragment"
     misc_feature    7404..10043
                     /note="assembly_fragment"
     misc_feature    10144..14123
                     /note="assembly_fragment"
     misc_feature    14224..18507
                     /note="assembly_fragment"
     misc_feature    18608..21538
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    21639..27268
                     /note="assembly_fragment"
     misc_feature    27369..33110
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
     misc_feature    33211..39973
                     /note="assembly_fragment"
     misc_feature    40074..45638
                     /note="assembly_fragment"
     misc_feature    45739..55034
                     /note="assembly_fragment"
     misc_feature    55135..62645
                     /note="assembly_fragment
     misc_feature    62746..77397
                     /note="assembly_fragment"
     misc_feature    77498..93454
                     /note="assembly_fragment
```

```
     misc_feature     /note="assembly_fragment"
                      93555..110875
     misc_feature     /note="assembly_fragment"
                      110976..128888
     misc_feature     /note="assembly_fragment"
                      128989..183298
                      /note="assembly_fragment"
ORIGIN

Query Match      87.6%;  Score 18.4;  DB 2;  Length 183298;
Best Local Similarity  95.0%;  Pred. No. 39;
Matches  19;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy     1 GGGATTCCTAGAGGGAAGGA 20
         | ||||||||||||||||||
Db 26670 GAGATTCCTAGAGGGAAGGA 26651


RESULT 12
AP001541
AP001541              212884 bp    DNA         linear   PRI 15-MAR-2003
LOCUS       Homo sapiens genomic DNA, chromosome 11q, clone:RP11-68304,
DEFINITION  complete sequence.
ACCESSION   AP001541
VERSION     AP001541.4   GI:13429924
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Yada,T., Totoki,Y., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2000)
REFERENCE   2  (bases 1 to 212884)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Yada,T., Totoki,Y., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2000) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Mar 21, 2001 this sequence version replaced gi:10130042.
FEATURES             Location/Qualifiers
     source          1..212884
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-68304"
ORIGIN

Query Match      87.6%;  Score 18.4;  DB 9;  Length 212884;
Best Local Similarity  95.0%;  Pred. No. 38;
Matches  19;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy     1 GGGATTCCTAGAGGGAAGGA 20
         | |||||||||||||||||
Db 91524 GAGATTCCTAGAGGGAAGGA 91543


RESULT 13
AC106477
AC106477             245820 bp    DNA         linear   HTG 13-MAY-2003
LOCUS       Rattus norvegicus clone CH230-165F14, *** SEQUENCING IN PROGRESS
DEFINITION  ***, 3 unordered pieces.
ACCESSION   AC106477
VERSION     AC106477.5   GI:30580899
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
```

```
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 245820)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji.M., Baca,B., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto.M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,K., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louleseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martin,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olarnpunsagoon,A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,M., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausen,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 245820)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 245820)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-MAY-2003) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On May 13, 2003 this sequence version replaced gi:24819137.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/).Each contig described
            in the feature table below represents a scaffold in the Atlas
```

assembly (a 'contig-scaffold'). Within each contig-scaffold,
individual sequence contigs are ordered and oriented, and separated
by sized gaps filled with Ns to the estimated size. The sequence
may extend beyond the ends of the contig and there may be sequence
contigs within a contig-scaffold that consist entirely of whole
genome shotgun sequence reads. Both end sequences and whole genome
shotgun sequence only contigs will be indicated in the feature
table.
---------- Genome Center
        Center: Baylor College of Medicine
        Center code: BCM
        Web site: http://www.hgsc.bcm.tmc.edu/
        Contact: hgsc-help@bcm.tmc.edu
---------- Project Information
        Center project name: GKZI
        Center clone name: CH230-165F14
---------- Summary Statistics
        Assembly program: Atlas 3.0;
        Consensus quality: 225710 bases at least Q40
        Consensus quality: 229029 bases at least Q30
        Consensus quality: 231433 bases at least Q20
        Estimated insert size: 235116; sum-of-contigs estimation
        Quality coverage: 6x in Q20 bases; sum-of-contigs estimation
----
* NOTE: Estimated insert size may differ from sequence length
*   (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
*   consists of 3 contigs. The true order of the pieces
*   is not known and their order in this sequence record is
*   arbitrary. Gaps between the contigs are represented as
*   runs of N, but the exact sizes of the gaps are unknown.
*   This record will be updated with the finished sequence
*   as soon as it is available and the accession number will
*   be preserved.
*   1 242644: contig of 242644 bp in length
*   242645 242744: gap of unknown length
*   242745 244368: contig of 1624 bp in length
*   244369 244468: gap of unknown length
*   244469 245820: contig of 1352 bp in length.

FEATURES   Location/Qualifiers
     source   1..245820
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-165F14"
     misc_feature   1..1822
                     /note="wgs_end_extension
                     clone_end:T7"
     misc_feature   2293..2621
                     /note="clone_boundary
                     clone_end:T7
                     site:EcoRI
                     end_sequence:BH273380"
     misc_feature   241929..242240
                     /note="clone_boundary
                     clone_end:Sp6
                     site:EcoRI
                     end_sequence:BH273381"
ORIGIN

Query Match   87.6%; Score 18.4; DB 2; Length 245820;
Best Local Similarity   95.0%; Pred. No. 38;
Matches   19; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy   1 GGGATTCCTAGAGGAAGGA 20
          |||||||||||| ||||||
Db 26316 GGGATTCCTAGAGGGTAGGA 26335

RESULT 14
AC016931
LOCUS   AC016931   169032 bp   DNA   linear   PRI 24-SEP-2002
DEFINITION Homo sapiens 3 BAC RP11-22E12 (Roswell Park Cancer Institute Human

ACCESSION   BAC Library) complete sequence.
            AC016931
VERSION   AC016931.21 GI:21617624
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 169032)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J, Draper,H., Dugan-Rocha,S., Durbin,K.J,
            Barnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., He,X.,
            Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C.,
            Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,
            Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
            Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J.,
            Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
            Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
            Loulseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
            Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
            Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P., Meador,M.,
            Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
            Moore,J., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,
            Nelson,D., Newton,J., Newtson,N., Nguyen,A., Nguyen,N., Nguyen,N.,
            Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G., Oragunye,N.,
            Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,
            Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M.,
            Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G., Scherer,S.,
            Scott,G., Shen,H., Shooshtari,N., Sisson,I., Sodergren,E.,
            Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,
            Tabor,P., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,
            Taylor,T., Telfrod,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,
            Vera,V., Villalon,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,
            Warren,R., Washington,C., Watlington,S., Williams,G.,
            Williamson,A., Wleczyk,R., Wooden,S., Worley,S., Worley,W., Wu,C., Wu,Y.,
            Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstock,G. and
            Gibbs,R.
  TITLE   Direct Submission
  JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 169032)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-DEC-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 169032)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (23-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4 (bases 1 to 169032)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission
  JOURNAL   Submitted (27-JUN-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5 (bases 1 to 169032)
  AUTHORS   Worley,K.C.
  TITLE   Direct Submission

```
JOURNAL     Submitted (03-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   6  (bases 1 to 169032)
AUTHORS     Worley,K.C.
TITLE       Direct Submission
JOURNAL     Submitted (24-SEP-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Jun 27, 2002 this sequence version replaced gi:21539051.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
              STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
              Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
              Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE:Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://www.hgsc.bcm.tmc.edu:8088/quality.info/genbank.annotation.ht
            ml.

FEATURES             Location/Qualifiers
     source          1..169032
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="3"
                     /clone="RP11-22E12"
     misc_feature    1..2004
                     /note="overlaps bases 84111..86114 of clone AC109912"
                     /function="clone overlap"
     STS             132..439
                     /standard_name="SHGC-77924"
     repeat_region   713..945
                     /rpt_family="MIR"
     repeat_region   895..954
                     /rpt_family="L2"
     repeat_region   complement(1740..2037)
                     /rpt_family="AluY"
     repeat_region   4358..4478
                     /rpt_family="L2"
     STS             4557..4888
                     /standard_name="SHGC-112333"
     repeat_region   9998..10146
                     /rpt_family="MIR"
     repeat_region   10551..10584
                     /rpt_family="(CA)n"
     repeat_region   complement(11053..11233)
                     /rpt_family="MER20"
     repeat_region   12536..12660
                     /rpt_family="L2"
     repeat_region   12891..13044
                     /rpt_family="MIR"
     repeat_region   complement(13402..13553)
                     /rpt_family="MIR"
     repeat_region   complement(14945..15295)
                     /rpt_family="THE1A"
     repeat_region   15305..15341
                     /rpt_family="AT_rich"
     repeat_region   complement(15453..15742)
                     /rpt_family="L1ME"
     repeat_region   complement(15798..16108)
                     /rpt_family="AluY"
     repeat_region   16184..16215
                     /rpt_family="(TG)n"
     repeat_region   16373..16479
                     /rpt_family="L2"
     repeat_region   complement(17305..17404)
                     /rpt_family="MIR"
     repeat_region   complement(18296..18406)
                     /rpt_family="MIR"
     repeat_region   complement(18489..18604)
                     /rpt_family="L2"
     repeat_region   19142..19175
                     /rpt_family="(TCCA)n"
     repeat_region   19607..19698
                     /rpt_family="GA-rich"
     repeat_region   19699..19908
                     /rpt_family="MLT1A1"
     repeat_region   19909..20212
                     /rpt_family="AluSx"
     repeat_region   20213..20377
                     /rpt_family="MLT1A1"
     repeat_region   20378..20396
                     /rpt_family="GA-rich"
     repeat_region   complement(20398..20526)
                     /rpt_family="MIR"
     repeat_region   22584..22614
                     /rpt_family="(TG)n"
     repeat_region   22625..22723
                     /rpt_family="MIR"
     repeat_region   23544..23565
                     /rpt_family="(CA)n"
     repeat_region   23566..23616
                     /rpt_family="(CA)n"
     repeat_region   23930..23977
                     /rpt_family="(GA)n"
     repeat_region   complement(24171..24302)
                     /rpt_family="MER5A"

Query Match          85.7%;  Score 18;  DB 9;  Length 169032;
Best Local Similarity 100.0%;  Pred. No. 65;
Matches  18;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy         4 ATTCCTAGAGGGAAGGAG 21
             |||||||||||||||||||
Db    131381 ATTCCTAGAGGGGAAGGAG 131398

RESULT 15
AC137676/c
LOCUS       AC137676           187214 bp   DNA   linear   HTG 28-FEB-2003
DEFINITION  Mus musculus clone RP23-1N19, WORKING DRAFT SEQUENCE, 9 unordered
            pieces.
ACCESSION   AC137676
VERSION     AC137676.3  GI:28604154
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
                  Quality coverage: 10.9 in Q20 bases; sum-of-contigs
------------------
* NOTE: This is a 'working draft' sequence. It currently
* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
*          1      3471: contig of 3471 bp in length
*       3472      3571: gap of 100 bp
*       3572      4642: contig of 1071 bp in length
*       4643      4742: gap of 100 bp
*       4743      6413: contig of 1671 bp in length
*       6414      6513: gap of 100 bp
*       6514     11072: contig of 4559 bp in length
*      11073     11172: gap of 100 bp
*      11173     14809: contig of 3637 bp in length
*      14810     14909: gap of 100 bp
*      14910     39178: contig of 24269 bp in length
*      39179     39278: gap of 100 bp
*      39279    105215: contig of 65937 bp in length
*      105216   105315: gap of 100 bp
*      105316   154654: contig of 49339 bp in length
*      154655   154754: gap of 100 bp
*      154755   187214: contig of 32460 bp in length.

FEATURES             Location/Qualifiers
     source          1..187214
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="RP23-1N19"
                     /clone_lib="RPCI-23 Female Mouse BAC"
     misc_feature    1..3471
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    3572..4642
                     /note="assembly_fragment"
     misc_feature    4743..6413
                     /note="assembly_fragment"
     misc_feature    6514..11072
                     /note="assembly_fragment"
     misc_feature    11173..14809
                     /note="assembly_fragment"
     misc_feature    14910..39178
                     /note="assembly_fragment"
     misc_feature    39279..105215
                     /note="assembly_fragment"
     misc_feature    105316..154654
                     /note="assembly_fragment"
     misc_feature    154755..187214
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:right"
ORIGIN

Query Match            85.7%;  Score 18;  DB 2;  Length 187214;
Best Local Similarity 100.0%;  Pred. No. 64;
Matches   18;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy        1 GGGATTCCTAGAGGGAAG 18
            ||||||||||||||||||
Db    37785 GGGATTCCTAGAGGGAAG 37768

Search completed: March 25, 2004, 12:49:59
Job time : 148.084 secs
```

REFERENCE
AUTHORS
TITLE
JOURNAL
REFERENCE
AUTHORS

Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 187214)
Birren,B., Nusbaum,C. and Lander,E.
Mus musculus, clone RP23-1N19
Unpublished
2 (bases 1 to 187214)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (27-NOV-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 187214)
Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
Collymore,A., Cook,A., Cooke,P., Corum.B., DeArellano,K.,
Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (28-FEB-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Feb 28, 2003 this sequence version replaced gi:28195913.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE
JOURNAL

REFERENCE
AUTHORS

TITLE
JOURNAL

COMMENT

------------------- Genome Center
          Center: Whitehead Institute/ MIT Center for Genome Research
          Center code: WIBR
          Web site: http://www-seq.wi.mit.edu
          Contact: sequence_submissions@genome.wi.mit.edu
------------------- Project Information
          Center project name: L26269
          Center clone name: 1_N19
------------------- Summary Statistics
          Sequencing vector: plasmid; n/a; 100% of reads
          Chemistry: Dye-terminator Big Dye; 100% of reads
          Assembly program: Phrap; version 0.960731
          Consensus quality: 184963 bases at least Q40
          Consensus quality: 185729 bases at least Q30
          Consensus quality: 186054 bases at least Q20
          Insert size: 180000; agarose-fp
          Insert size: 186414; sum-of-contigs
          Quality coverage: 11.3 in Q20 bases; agarose-fp

```
            GenCore version 5.1.6
     Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:    March 25, 2004, 07:40:23 ; Search time 21.7931 Seconds
                                       (without alignments)
                                       4093.601 Million cell updates/sec

Title:            US-09-963-285-1_COPY_403_423
Perfect score:    21
Sequence:         1 gggattcctagagggaaggag 21

Scoring table:    IDENTITY_NUC
                  Gapop 10.0 , Gapext 1.0

Searched:         3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:    6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :    N_Geneseq_29Jan04:*
              1: geneseqn1980s:*
              2: geneseqn1990s:*
              3: geneseqn2000s:*
              4: geneseqn2001as:*
              5: geneseqn2001bs:*
              6: geneseqn2002s:*
              7: geneseqn2003as:*
              8: geneseqn2003bs:*
              9: geneseqn2003cs:*
             10: geneseqn2004s:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Match	Query Match Length	DB	ID	Description
1		21	100.0	6021	6	ABK48986	Abk48986 Genomic D
2		21	100.0	6458	6	ABK48984	Abk48984 Genomic D
3		19.4	92.4	8865	4	AAK87029	Aak87029 Human imm
4		17.8	84.8	9951	6	ABK39928	Abk39928 Human che
5		17.4	82.9	549	7	ABX73141	Abx73141 Metabolic
6	c	16.8	80.0	2796	3	AAA28713	Aaa28713 C. pneumo
7	c	16.8	80.0	2817	3	ABL91186	Abl91186 Chlamydia
8	c	16.8	80.0	3000	3	AAA28712	Aaa28712 C. pneumo
9	c	16.8	80.0	32177	4	AAS29828	Aas29828 Human cyt
10	c	16.8	80.0	32177	5	AAL05374	Aal05374 Human rep
11	c	16.8	80.0	32177	5	ABA18493	Aba18493 Human ner
12	c	16.8	80.0	110000	2	AAX91990_05	Continuation (6 of
13	c	16.8	80.0	273254	4	AAC81914	Aac81914 Chlamydia
14		16.8	80.0	325791	4	AAS43104	Aas43104 Human Oes
15	c	16.4	78.1	303	6	AAC24739	Aac24739 Human sec
16		16.4	78.1	307	6	ABN60795	Abn60795 Human can
17		16.4	78.1	422	8	ACH21145	Ach21145 Human adu
18	c	16.4	78.1	436	6	ABV97368	Abv97368 Human pan
19	c	16.4	78.1	591	2	AAX97629	Aax97629 Extended
20		16.4	78.1	610	9	ADB55830	Adb55830 Toxicity-
21	c	16.4	78.1	1358	6	AAS62425	Aas62425 cDNA sequ
22	c	16.4	78.1	2625	4	AAH14444	Aah14444 Human cDN
23	c	16.4	78.1	2805	6	ABL91187	Abl91187 Chlamydia
24	c	16.4	78.1	3003	3	AAA28711	Aaa28711 C. pneumo
25	c	16.4	78.1	3200	3	AAA28710	Aaa28710 C. pneumo
26		16.4	78.1	29956	8	ADA02963	Ada02963 Mouse Lck
27		16.4	78.1	29956	9	ADB72701	Adb72701 Mouse Lck
28		16.4	78.1	29956	9	ADC85443	Adc85443 Mouse Lck
29		16.4	78.1	31749	4	AAK72959	Aak72959 Human imm
30	c	16.4	78.1	78785	7	AAL60948	Aal60948 Human nep
31		16.4	78.1	78925	7	AAC89888	Aac89888 Human FN
32		16.4	78.1	110000	7	ACF42745_1	Continuation (2 of
33	c	16.4	78.1	202001	6	ABS52506	Abs52506 Human tra
34		16.2	77.1	332	6	ABS72685	Abs72685 Human gen
35	c	16.2	77.1	384	6	ABN17597	Abn17597 Human ORF
36	c	16.2	77.1	479	8	ACH44328	Ach44328 Human foe
37		16.2	77.1	520	4	AAL16066	Aal16066 Human bre
38	c	16.2	77.1	600	5	ABV54015	Abv54015 Human pro
39		16.2	77.1	900	5	AAF92506	Aaf92506 Human T2R
40		16.2	77.1	900	5	AAS70020	Aas70020 DNA encod
41	c	16.2	77.1	3117	5	ABA15179	Aba15179 Human ner
42		16.2	77.1	3117	5	ABA15177	Aba15177 Human ner
43	c	16.2	77.1	3117	5	ABA15180	Aba15180 Human ner
44	c	16.2	77.1	3117	5	ABA15178	Aba15178 Human ner
45	c	16.2	77.1	3172	6	AAS62320	Aas62320 cDNA sequ

ALIGNMENTS

```
RESULT 1
ABK48986
ID   ABK48986 standard; DNA; 6021 BP.
XX
AC   ABK48986;
XX
DT   02-JUL-2002  (first entry)
XX
DE   Genomic DNA encoding mouse mesenchyme forkhead-1 (MHF-1)/FOXC2.
XX
KW   Transcription factor; FOXC2; antidiabetic; anorectic; antilipaemic;
KW   cardiovascular; FOXC2 polypeptide expression modulator; FOXC2 promoter;
KW   diabetes mellitus; obesity; hypercholesterolaemia; dyslipidaemia;
KW   cardiovascular disease; mouse medenchyme forkhead 1; MHF-1; gene; ds.
OS   Mus musculus.
XX
FH   Key            Location/Qualifiers
FT   CDS            2070..3554
FT                  /*tag= a
FT                  /product= "MHF-1 or FOXC2"
FT                  /note= "Mouse mesenchyme forkhead 1"
XX
PN   WO200227008-A1.
XX
PD   04-APR-2002.
XX
PF   26-SEP-2001; 2001WO-SE002098.
XX
PR   26-SEP-2000; 2000SE-00003435.
PR   10-OCT-2000; 2000US-0238897P.
PR   09-NOV-2000; 2000SE-00004102.
XX
PA   (BIOV-) BIOVITRUM AB.
XX
PI   Enerbaeck S, Krook K, Rondahl L, Wasserman WW;
XX
DR   WPI; 2002-352129/38.
DR   P-PSDB; AAU79818.
XX
PT   An isolated FOXC2 promoter region that modulates the expression of a
PT   FOXC2 polypeptide is useful for treating obesity and type II diabetes
PT   mellitus.
XX
PS   Claim 22; Page 45-46; 62pp; English.
XX
```

```
CC   The invention describes an isolated human FOXC2 promoter region. The
CC   mammalian FOXC2 promoter region for the modulation of FOXC2 polypeptide
CC   expression has therapeutic value in treating type II diabetes mellitus,
CC   obesity, hypercholesterolaemia, other cardiovascular diseases or
CC   dyslipidaemias. This sequence encodes the mouse mesenchyme forkhead 1
CC   (MHF-1) protein (also called FOXC2 transcription factor), described in
CC   the method of the invention
XX
SQ   Sequence 6021 BP; 1350 A; 1675 C; 1642 G; 1354 T; 0 U; 0 Other;

Query Match             100.0%;  Score 21;  DB 6;  Length 6021;
Best Local Similarity   100.0%;  Pred. No. 2.5;
Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGATTCCTAGAGGGAAGGAG 21
          |||||||||||||||||||||
Db    385 GGGATTCCTAGAGGGAAGGAG 405

RESULT 2
ABK48984
ID   ABK48984 standard; DNA; 6458 BP.
XX
AC   ABK48984;
XX
DT   02-JUL-2002  (first entry)
XX
DE   Genomic DNA encoding human transcription factor FOXC2.
XX
KW   Transcription factor; FOXC2; antidiabetic; anorectic; antilipaemic;
KW   cardiovascular; FOXC2 polypeptide expression modulator; FOXC2 promoter;
KW   diabetes mellitus; obesity; hypercholesterolaemia; dyslipidaemia;
KW   cardiovascular disease; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   exon            1..215
FT                   /*tag= b
FT                   /note= "First exon according to the alternative
FT                   transcript"
FT   5'UTR           1..186
FT                   /*tag= a
FT   misc_feature    187..215
FT                   /*tag= c
FT                   /note= "Region coding for 5'part of alternative protein"
FT   misc_signal     215..216
FT                   /*tag= d
FT                   /note= "Alternative first exon splice site"
FT   misc_difference 216..475
FT                   /*tag= e
FT                   /note= "First exon according to the alternative
FT                   transcript"
FT   misc_difference 223..231
FT                   /*tag= f
FT                   /note= "Fragment of the FOXC2 enhancer. Specifically
FT                   claimed in claim 16"
FT   misc_difference 359..375
FT                   /*tag= g
FT                   /note= "Fragment of the FOXC2 enhancer. Specifically
FT                   claimed in claim 12"
FT   misc_difference 378..402
FT                   /*tag= h
FT                   /note= "Fragment of the FOXC2 enhancer. Specifically
FT                   claimed in claim 13"
FT   misc_difference 403..423
FT                   /*tag= i
FT                   /note= "Fragment of the FOXC2 enhancer. Specifically
FT                   claimed in claim 14"
FT   misc_difference 1250..2235
FT                   /*tag= k
FT                   /note= "Fragment of the FOXC2 promoter. Specifically
FT                   claimed in claim 3"
FT   misc_difference 1250..1749
FT                   /*tag= j
FT                   /note= "Fragment of the FOXC2 promoter. Specifically
FT                   claimed in claim 2"
FT   misc_difference 1692..1703
FT                   /*tag= l
FT                   /note= "Fragment of the FOXC2 promoter. Specifically
FT                   claimed in claim 1"
FT   exon            1746..4629
FT                   /*tag= n
FT                   /note= "First exon according to the published form of the
FT                   transcript"
FT   5'UTR           1746..2234
FT                   /*tag= m
FT   CDS             2235..3740
FT                   /*tag= o
FT                   /product= "FOXC2"
FT                   /note= "Transcription factor"
FT   misc_binding    2448..2735
FT                   /*tag= p
FT                   /note= "Region coding for DNA-binding domain"
FT   exon            2516..4629
FT                   /*tag= r
FT                   /note= "Second exon according to the alternative
FT                   transcript"
FT   misc_feature    2516..3740
FT                   /*tag= q
FT                   /note= "Portion of polypeptide used in alternative
FT                   transcript"
FT   3'UTR           3741..4629
FT                   /*tag= s
XX
PN   WO20227008-A1.
XX
PD   04-APR-2002.
XX
PF   26-SEP-2001; 2001WO-SE002098.
XX
PR   26-SEP-2000; 2000SE-00003435.
PR   10-OCT-2000; 2000US-0238897P.
PR   09-NOV-2000; 2000SE-00004102.
XX
PA   (BIOV-) BIOVITRUM AB.
XX
PI   Enerbaeck S,  Krook K,  Rondahl L,  Wasserman WW;
XX
DR   WPI; 2002-352129/38.
DR   P-PSDB; AAU79816.
XX
PT   An isolated FOXC2 promoter region that modulates the expression of a
PT   FOXC2 polypeptide is useful for treating obesity and type II diabetes
PT   mellitus.
XX
PS   Claim 3; Page 33-38; 62pp; English.
XX
CC   The invention describes an isolated human FOXC2 promoter region. The
CC   mammalian FOXC2 promoter region for the modulation of FOXC2 polypeptide
CC   expression has therapeutic value in treating type II diabetes mellitus,
CC   obesity, hypercholesterolaemia, other cardiovascular diseases or
CC   dyslipidaemias. This sequence encodes the human transcription factor
CC   FOXC2, the promoter of which is described in the invention
XX
SQ   Sequence 6458 BP; 1443 A; 1826 C; 1715 G; 1474 T; 0 U; 0 Other;

Query Match             100.0%;  Score 21;  DB 6;  Length 6458;
Best Local Similarity   100.0%;  Pred. No. 2.5;
Matches   21;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy      1 GGGATTCCTAGAGGGAAGGAG 21
          |||||||||||||||||||||
Db    403 GGGATTCCTAGAGGGAAGGAG 423
```

```
RESULT 3
AAK87029
ID   AAK87029 standard; DNA; 8865 BP.
XX
AC   AAK87029;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:41841.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2001; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0209963P.
PR   26-JUL-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224518P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0252214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0262279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226866P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0270009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0232401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0234997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0235484P.
PR   27-SEP-2000; 2000US-0235834P.
PR   27-SEP-2000; 2000US-0235836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236367P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0209960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241882P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
PR   17-NOV-2000; 2000US-0249214P.
PR   17-NOV-2000; 2000US-0249215P.
PR   17-NOV-2000; 2000US-0249216P.
PR   17-NOV-2000; 2000US-0249217P.
PR   17-NOV-2000; 2000US-0249218P.
PR   17-NOV-2000; 2000US-0249244P.
PR   17-NOV-2000; 2000US-0249245P.
PR   17-NOV-2000; 2000US-0249264P.
PR   17-NOV-2000; 2000US-0249265P.
PR   17-NOV-2000; 2000US-0249297P.
PR   17-NOV-2000; 2000US-0249299P.
PR   17-NOV-2000; 2000US-0249300P.
PR   01-DEC-2000; 2000US-0250160P.
PR   01-DEC-2000; 2000US-0250391P.
```

PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251988P.
PR   05-DEC-2000;   2000US-0256719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0256678P.
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 41841; 3071pp + Sequence Listing; English.
XX
CC   AAK4951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM2170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention
XX
SQ   Sequence 8865 BP; 2486 A; 1904 C; 2073 G; 2401 T; 0 U; 1 Other;

Query Match   92.4%; Score 19.4; DB 4; Length 8865;
Best Local Similarity   95.2%; Pred. No. 15;
  Matches   20; Conservative   0; Mismatches   1; Indels   0; Gaps   0;

Qy   1 GGGATTCCTAGAGGGAAGGAG 21
              ||||||||||||| ||||||
Db   857 GGGATTCCTAGAGGTAAGGAG 877

RESULT 4
ABK39928
ID   ABK39928 standard; DNA; 9951 BP.
XX
AC   ABK39928;
XX
DT   21-MAY-2002 (first entry)
XX
DE   Human chemically pretreated gene sequence #5 strand 1.
XX
KW   Human; ds; bisulphite treatment; CpG; DNA methylation; cancer; tumour;
KW   cytostatic; ALDH6; CYP11A; CYP11B1; CYP3A3; DPYD; EPHX2; OCLN; TXNRD1;
KW   UGT8; MRP; pharmacogenomics; SNP; single nucleotide polymorphism.
XX
OS   Homo sapiens.
XX
PN   WO20020202806-A2.
XX
PD   10-JAN-2002.
XX
PF   29-JUN-2001; 2001WO-EP007470.

XX
PR   30-JUN-2000;   2000DE-01032529.
PR   01-SEP-2000;   2000DE-01043826.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K;
XX
DR   WPI; 2002-154757/20.
XX
PT   New nucleic acid, oligonucleotides and peptide nucleic acid-oligomers,
PT   useful for detecting cytosine methylation state of genes associated with
PT   pharmacogenomics and for therapy of diseases e.g. cancer.
XX
PS   Claim 1; SEQ ID NO 9; 24pp; English.
XX
CC   The invention relates to a nucleic acid comprising a sequence at least 18
CC   bases in length of a segment of the chemically pretreated DNA of genes
CC   associated with pharmacogenomics according to one of the sequences of the
CC   genes ALDH6 (NM_000693), CYP11A (NM_000781), CYP11B1 (NM_000497), CYP3A3
CC   (NM_000776 and NM_017460), DPYD (NM_000110), EPHX2 (NM_001979), OCLN
CC   (NM_002538), TXNRD1 (NM_003330), UGT8 (NM_003360), MRP (NM_004996),
CC   NM_019900, NM_019901, NM_019902, NM_019862, NM_019898, NM_019899) and
CC   their complementary sequences, or a sequence (S1) chosen from 87
CC   sequences and their complements. The chemical pretreatment is bisulphite
CC   treatment to convert cytosines (but not methyl-cytosines) into uracil.
CC   Also included are an oligomer (II) in particular an oligonucleotide or a
CC   peptide nucleic acid (PNA)-oligomer, comprising in each case at least one
CC   base sequence having a length of 9 nucleotides which hybridises to or is
CC   identical to a chemically pretreated DNA of genes associated with
CC   pharmacogenomics and their complements, arranged in an array for
CC   analysing diseases associated with the methylation state (CpG) and/or
CC   detecting SNPs (single nucleotide polymorphisms) of the 87 sequences. The
CC   oligomers may also be used as PCR primers. The set of 87 nucleic acids
CC   and their complements is useful for diagnosis and therapy of solid
CC   tumours and cancer. The present sequence represents one the 87 DNA
CC   sequences or its complement. Note:The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 9951 BP; 2117 A; 204 C; 2950 G; 4680 T; 0 U; 0 Other;

Query Match   84.8%; Score 17.8; DB 6; Length 9951;
Best Local Similarity   90.5%; Pred. No. 89;
  Matches   19; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 GGGATTCCTAGAGGGAAGGAG 21
              ||||||| |||||| ||||||
Db   9036 GGGATTCGGAGAGGGAAGGAG 9056

RESULT 5
ABX73141
ID   ABX73141 standard; cDNA; 549 BP.
XX
AC   ABX73141;
XX
DT   17-MAR-2003 (first entry)
XX
DE   Metabolic pathway (MP) protein cDNA #23.
XX
KW   Metabolic pathway protein; MP; gene; ss; moss; alga; Ceratodon purpureus;
KW   fine chemical; Corynebacterium; Brevibacterium; vitamin; cofactor;
KW   metabolism; essential amino acid; tocopherol; riboflavin; vitamin C.
XX
OS   Physcomitrella patens.
XX
PN   US2002142422-A1.
XX
PD   03-OCT-2002.
XX
PF   12-DEC-2000; 2000US-00734017.

XX
PR   16-DEC-1999;   99US-0171100P.
XX
PA   (LERC/) LERCHL J.
PA   (RENZ/) RENZ A.
PA   (EHRH/) EHRHARDT T.
PA   (REIN/) REINDL A.
PA   (CIRP/) CIRPUS P.
PA   (BISC/) BISCHOFF F.
PA   (FRAN/) FRANK M.
PA   (FREU/) FREUND A.
PA   (DUWE/) DUWENIG E.
PA   (SCHM/) SCHMIDT R.
PA   (RESK/) RESKI R.
XX
PI   Lerchl J,   Renz A,   Ehrhardt T,   Reindl A,   Cirpus P,   Bischoff F;
PI   Frank M,   Freund A,   Duwenig E,   Schmidt R,   Reski R;
XX
DR   WPI; 2003-155946/15.
DR   P-PSDB; ABU54893.
XX
PT   Novel isolated nucleic acids from moss e.g., Physcomitrella patens,
PT   encoding a metabolic pathway protein, useful for modifying production of
PT   amino acids, vitamins, cofactors in plants, algae and microorganisms.
XX
PS   Disclosure; Page 36; 107pp; English.
XX
CC   The invention relates to a nucleic acid molecule from a moss (e.g.,
CC   Physcomitrella patens or Ceratodon purpureus) encoding a metabolic
CC   pathway (MP) protein or its portion. A vector comprising a nucleic acid
CC   is useful for producing a fine chemical which involves culturing a cell
CC   containing the vector such that the fine chemical is produced and then
CC   recovered. The cell is a microorganism belonging to the genus
CC   Corynebacterium or Brevibacterium, to the genus mosses or algae, or is a
CC   plant cell. The expression of the nucleic acid from the vector in the
CC   cell results in modulation of the production of the fine chemical such as
CC   amino acids, vitamins, cofactors, nucleotides and/or nucleosides. The MP
CC   nucleic acid molecules are useful for modulating production of fine
CC   chemicals in microorganisms, algae and plants, either directly or
CC   indirectly. Plant genes originating from P. patens are useful for
CC   modifying metabolism of essential amino acids, tocopherol, riboflavin,
CC   vitamin C, etc., in plants as well as algae and microorganisms, enabling
CC   these host cells to increase their capacity to produce the respective
CC   components as well as improving survival and fitness of the cell. The
CC   nucleic acids are useful for identifying an organism as being P. patens
CC   or its close relative in a mixed population of microorganisms, as markers
CC   for specific regions of the genome and for evolutionary and protein
CC   structural studies. Sequences ABX73119-ABX73160 represent cDNA molecules
CC   encoding MP proteins of the invention
XX
SQ   Sequence 549 BP; 120 A; 136 C; 148 G; 145 T; 0 U; 0 Other;

Query Match   82.9%;   Score 17.4;   DB 7;   Length 549;
Best Local Similarity 94.7%;   Pred. No. 1.2e+02;
Matches   18;   Conservative   0;   Mismatches   1;   Indels   0;   Gaps   0;

Qy   1 GGGATTCCTAGAGGGAAGG 19
         ||||||||| ||||||||||
Db   247 GGGATTCCTCCTGGAGGGAAGG 265

RESULT 6
AAA28713/c
ID   AAA28713 standard; DNA; 2796 BP.
XX
AC   AAA28713;
XX
DT   15-SEP-2003 (revised)
DT   29-AUG-2000 (first entry)
XX
DE   C. pneumoniae CPN100628 open reading frame.
XX
KW   Antigen; anti-inflammatory; respiratory; antibacterial; anti-asthmatic;

KW   anti-arteriosclerotic; vaccine; ds.
XX
OS   Chlamydophila pneumoniae.
XX
PN   WO200024765-A2.
XX
PD   04-MAY-2000.
XX
PF   28-OCT-1999;   99WO-CA000992.
XX
PR   28-OCT-1998;   98US-0106034P.
PR   28-OCT-1998;   98US-0106039P.
PR   28-OCT-1998;   98US-0106042P.
PR   28-OCT-1998;   98US-0106044P.
PR   29-OCT-1998;   98US-0106072P.
PR   29-OCT-1998;   98US-0106073P.
PR   29-OCT-1998;   98US-0106074P.
PR   29-OCT-1998;   98US-0106087P.
PR   02-NOV-1998;   98US-0105587P.
PR   02-NOV-1998;   98US-0105588P.
PR   02-NOV-1998;   98US-0105589P.
PR   02-NOV-1998;   98US-0107034P.
PR   02-NOV-1998;   98US-0107035P.
XX
PA   (CONN-) CONNAUGHT LAB LTD.
XX
PI   Murdin AD,   Oomen RP,   Wang J;
XX
DR   WPI; 2000-350688/30.
DR   P-PSDB; AAY92833.
XX
PT   Chlamydia antigenes and the proteins they encode, useful for vaccinating
PT   against Chlamydia infections that affect the respiratory tract.
XX
PS   Claim 2; Fig 23; 226pp; English.
XX
CC   The nucleic acids may be used for the recombinant production of the
CC   chlamydia polypeptides (either in vivo or in vitro) according to standard
CC   recombinant DNA methodologies. The polypeptides may then be used to
CC   vaccinate against Chlamydia infections in mammals. Chlamydia, such as C.
CC   pneumoniae, are pathogens responsible for upper respiratory tract
CC   infections such as community acquired pneumonia, acute respiratory
CC   disease and bronchitis and may be implicated in atherosclerotic changes
CC   and asthma. The nucleic acids may also be used as probes for detecting
CC   the presence of Chlamydia nucleic acids in samples (and therefore
CC   diagnose infections) and the proteins may be used as antigens for the
CC   production of antibodies that may be used to detect Chlamydia proteins in
CC   samples (e.g. via enzyme linked immunosorbant assay (ELISA)). (Updated on
CC   15-SEP-2003 to standardise OS field)
XX
SQ   Sequence 2796 BP; 798 A; 703 C; 528 G; 767 T; 0 U; 0 Other;

Query Match   80.0%;   Score 16.8;   DB 3;   Length 2796;
Best Local Similarity 90.0%;   Pred. No. 2.5e+02;
Matches   18;   Conservative   0;   Mismatches   2;   Indels   0;   Gaps   0;

Qy   2 GGATTCCTAGAGGGAAGGAG 21
         ||||||||| |||| |||||
Db 2509 GGATTCCTACAGGGAGGGAG 2490

RESULT 7
ABL91186/c
ID   ABL91186 standard; DNA; 2817 BP.
XX
AC   ABL91186;
XX
DT   29-AUG-2003 (revised)
DT   29-JUL-2002 (first entry)
XX
DE   Chlamydia pneumoniae cp6751 ORF DNA, SEQ ID NO:6.
XX
KW   Chlamydial infection; antigen; immunogen; vaccine; diagnosis;

KW human respiratory disease; cardiovascular disease; atherosclerosis;
KW coronary artery disease; carotid artery stenosis; myocardial infarction;
KW cerebrovascular disease; aortic aneurysm; claudication; stroke;
KW strain CWL029; open reading frame; ORF; gene; ds.
XX
OS Chlamydophila pneumoniae.
XX
FH Key   Location/Qualifiers
FT CDS   1..2817
FT   /*tag= a
FT   /product= "cp6751"
FT sig_peptide   1..51
FT   /*tag= b
FT mat_peptide   52..2814
FT   /*tag= c
FT   /product= "Mature protein"
XX
PN WO200202606-A2.
XX
PD 10-JAN-2002.
XX
PF 03-JUL-2001; 2001WO-IB001445.
XX
PR 03-JUL-2000; 2000GB-00016363.
PR 11-JUL-2000; 2000GB-00017047.
PR 21-JUL-2000; 2000GB-00017983.
PR 07-AUG-2000; 2000GB-00019368.
PR 18-AUG-2000; 2000GB-00020440.
PR 14-SEP-2000; 2000GB-00022583.
PR 10-NOV-2000; 2000GB-00027549.
PR 22-DEC-2000; 2000GB-00031706.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Ratti G, Grandi G;
XX
DR WPI; 2002-154726/20.
DR N-PSDB; ABB90528.
XX
PT Novel Chlamydia pneumoniae protein useful in the manufacture of a
PT medicament for treatment or prevention of infection due to Chlamydia,
PT preferably Chlamydia pneumoniae, and for diagnostic purposes.
XX
PS Claim 5; Page 43-44; 364pp; English.
XX
CC Sequences ABB90526-ABB90715 represent novel proteins from Chlamydia
CC pneumoniae (strain CWL029), and ABL91184-ABL91373 represent DNA encoding
CC them. The proteins are predicted to be immunogenic and may therefore be
CC useful in vaccine production and for diagnostic purposes. Chlamydia
CC pneumoniae is a common cause of respiratory disease in humans, and is
CC also involved in the development of cardiovascular diseases such as
CC atherosclerosis, coronary artery disease, carotid artery stenosis,
CC myocardial infarction, cerebrovascular disease, aortic aneurysm,
CC claudication and stroke. The proteins and nucleic acids of the invention
CC may be used in vaccines and pharmaceutical compositions for the
CC prevention or treatment of chlamydial infections, particularly Chlamydia
CC pneumoniae infections. The proteins may also be used in the detection of
CC Chlamydia pneumoniae, and the nucleic acids may be used in PCR, branched
CC DNA probe assay or blotting techniques for determining Chlamydia
CC pneumoniae gene expression. The present sequence represents a
CC specifically claimed DNA which encodes a Chlamydia pneumoniae protein of
CC the invention. (Updated on 29-AUG-2003 to standardise OS field)
XX
SQ Sequence 2817 BP; 800 A; 706 C; 534 G; 777 T; 0 U; 0 Other;

Query Match   80.0%; Score 16.8; Pred. No. 2.5e+02; Length 2817;
Best Local Similarity 90.0%;
Matches   18; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   2 GGATTCCTAGAGGGAAGGAG 21
          |||||| |||||||||||||
Db   2530 GGATTCCTACAGGGAGGGAG 2511

RESULT 8
AAA28712/c
ID AAA28712 standard; DNA; 3000 BP.
XX
AC AAA28712;
XX
DT 15-SEP-2003 (revised)
DT 29-AUG-2000 (first entry)
XX
DE C. pneumoniae CPN100628 gene.
XX
KW Antigen; anti-inflammatory; respiratory; antibacterial; anti-asthmatic;
KW anti-arteriosclerotic; vaccine; ds.
XX
OS Chlamydophila pneumoniae.
XX
FH Key   Location/Qualifiers
FT CDS   101..2896
FT   /*tag= a
XX
PN WO200024765-A2.
XX
PD 04-MAY-2000.
XX
PF 28-OCT-1999; 99WO-CA000992.
XX
PR 28-OCT-1998; 98US-0106034P.
PR 28-OCT-1998; 98US-0106039P.
PR 28-OCT-1998; 98US-0106042P.
PR 28-OCT-1998; 98US-0106044P.
PR 29-OCT-1998; 98US-0106072P.
PR 29-OCT-1998; 98US-0106073P.
PR 29-OCT-1998; 98US-0106074P.
PR 29-OCT-1998; 98US-0106087P.
PR 02-NOV-1998; 98US-0105587P.
PR 02-NOV-1998; 98US-0105588P.
PR 02-NOV-1998; 98US-0105589P.
PR 02-NOV-1998; 98US-0107034P.
PR 02-NOV-1998; 98US-0107035P.
XX
PA (CONN-) CONNAUGHT LAB LTD.
XX
PI Murdin AD, Oomen RP, Wang J;
XX
DR WPI; 2000-350688/30.
DR P-PSDB; AAY92833.
XX
PT Chlamydia antigenes and the proteins they encode, useful for vaccinating
PT against Chlamydia infections that affect the respiratory tract.
XX
PS Claim 2; Fig 23; 226pp; English.
XX
CC The nucleic acids may be used for the recombinant production of the
CC Chlamydia polypeptides (either in vivo or in vitro) according to standard
CC recombinant DNA methodologies. The polypeptides may then be used to
CC vaccinate against Chlamydia infections in mammals. Chlamydia, such as C.
CC pneumoniae, are pathogens responsible for upper respiratory tract
CC infections such as community acquired pneumonia, acute respiratory
CC disease and bronchitis and may be implicated in atherosclerotic changes
CC and asthma. The nucleic acids may also be used as probes for detecting
CC the presence of Chlamydia nucleic acids in samples (and therefore
CC diagnose infections) and the proteins may be used as antigens for the
CC production of antibodies that may be used to detect Chlamydia proteins in
CC samples (e.g. via enzyme linked immunosorbant assay (ELISA)). (Updated on
CC 15-SEP-2003 to standardise OS field)
XX
SQ Sequence 3000 BP; 865 A; 739 C; 562 G; 834 T; 0 U; 0 Other;

Query Match   80.0%; Score 16.8; Score 16.8; DB 3; Length 3000;
Best Local Similarity 90.0%; Pred. No. 2.5e+02;
Matches   18; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

```
Qy            2  GGATTCCTAGAGGGAAGGAGGAG 21
                 ||||||||||| |||| ||||||
Db         2609  GGATTCCTACAGGGAGGGAG 2590

RESULT 9
AAS29828/c
ID   AAS29828 standard; DNA; 32177 BP.
XX
AC   AAS29828;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human cytoskeletal element-related polypeptide encoding genomic DNA #15.
XX
KW   Cytoskeletal element-related protein; human; mouse; rabbit; goat; horse;
KW   cat; dog; chicken; sheep; immunosuppressive; antiarthritic; vasotropic;
KW   antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;
KW   cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;
KW   ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;
KW   hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;
KW   cerebrovascular disorder; nervous system disorder; bacterial infection;
KW   fungal infection; viral infection; ocular disorder; endocrine disorder;
KW   gastrointestinal disorder; renal disorder; respiratory disorder;
KW   wound healing; skin aging; organ transplantation; food preservative;
KW   tissue regeneration; anti-infertility; food additive.
XX
OS   Homo sapiens.
XX
PN   WO200155168-A1.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001331.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189874P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
PR   30-JUN-2000; 2000US-0215135P.
PR   07-JUL-2000; 2000US-0216647P.
PR   07-JUL-2000; 2000US-0216880P.
PR   11-JUL-2000; 2000US-0217487P.
PR   11-JUL-2000; 2000US-0217496P.
PR   14-JUL-2000; 2000US-0218290P.
PR   26-JUL-2000; 2000US-0220963P.
PR   14-AUG-2000; 2000US-0220964P.
PR   14-AUG-2000; 2000US-0224418P.
PR   14-AUG-2000; 2000US-0224519P.
PR   14-AUG-2000; 2000US-0225213P.
PR   14-AUG-2000; 2000US-0225214P.
PR   14-AUG-2000; 2000US-0225266P.
PR   14-AUG-2000; 2000US-0225267P.
PR   14-AUG-2000; 2000US-0225268P.
PR   14-AUG-2000; 2000US-0225270P.
PR   14-AUG-2000; 2000US-0225447P.
PR   14-AUG-2000; 2000US-0225757P.
PR   14-AUG-2000; 2000US-0225758P.
PR   14-AUG-2000; 2000US-0225759P.
PR   18-AUG-2000; 2000US-0226279P.
PR   22-AUG-2000; 2000US-0226681P.
PR   22-AUG-2000; 2000US-0226868P.
PR   22-AUG-2000; 2000US-0227182P.
PR   23-AUG-2000; 2000US-0227009P.
PR   30-AUG-2000; 2000US-0228924P.
PR   01-SEP-2000; 2000US-0229287P.
PR   01-SEP-2000; 2000US-0229343P.
PR   01-SEP-2000; 2000US-0229344P.
PR   01-SEP-2000; 2000US-0229345P.
PR   05-SEP-2000; 2000US-0229509P.
PR   05-SEP-2000; 2000US-0229513P.
PR   06-SEP-2000; 2000US-0230437P.
PR   06-SEP-2000; 2000US-0230438P.
PR   08-SEP-2000; 2000US-0231242P.
PR   08-SEP-2000; 2000US-0231243P.
PR   08-SEP-2000; 2000US-0231244P.
PR   08-SEP-2000; 2000US-0231413P.
PR   08-SEP-2000; 2000US-0231414P.
PR   08-SEP-2000; 2000US-0232080P.
PR   08-SEP-2000; 2000US-0232081P.
PR   12-SEP-2000; 2000US-0231968P.
PR   14-SEP-2000; 2000US-0232397P.
PR   14-SEP-2000; 2000US-0232398P.
PR   14-SEP-2000; 2000US-0232399P.
PR   14-SEP-2000; 2000US-0232400P.
PR   14-SEP-2000; 2000US-0224401P.
PR   14-SEP-2000; 2000US-0233063P.
PR   14-SEP-2000; 2000US-0233064P.
PR   14-SEP-2000; 2000US-0233065P.
PR   21-SEP-2000; 2000US-0234223P.
PR   21-SEP-2000; 2000US-0234274P.
PR   25-SEP-2000; 2000US-0244997P.
PR   25-SEP-2000; 2000US-0234998P.
PR   26-SEP-2000; 2000US-0255484P.
PR   27-SEP-2000; 2000US-0255834P.
PR   27-SEP-2000; 2000US-0255836P.
PR   29-SEP-2000; 2000US-0236327P.
PR   29-SEP-2000; 2000US-0236377P.
PR   29-SEP-2000; 2000US-0236368P.
PR   29-SEP-2000; 2000US-0236369P.
PR   29-SEP-2000; 2000US-0236370P.
PR   02-OCT-2000; 2000US-0236802P.
PR   02-OCT-2000; 2000US-0237037P.
PR   02-OCT-2000; 2000US-0237038P.
PR   02-OCT-2000; 2000US-0237039P.
PR   02-OCT-2000; 2000US-0237040P.
PR   13-OCT-2000; 2000US-0239935P.
PR   13-OCT-2000; 2000US-0239937P.
PR   20-OCT-2000; 2000US-0240960P.
PR   20-OCT-2000; 2000US-0241221P.
PR   20-OCT-2000; 2000US-0241785P.
PR   20-OCT-2000; 2000US-0241786P.
PR   20-OCT-2000; 2000US-0241787P.
PR   20-OCT-2000; 2000US-0241808P.
PR   20-OCT-2000; 2000US-0241809P.
PR   20-OCT-2000; 2000US-0241826P.
PR   01-NOV-2000; 2000US-0244617P.
PR   08-NOV-2000; 2000US-0246474P.
PR   08-NOV-2000; 2000US-0246475P.
PR   08-NOV-2000; 2000US-0246476P.
PR   08-NOV-2000; 2000US-0246477P.
PR   08-NOV-2000; 2000US-0246478P.
PR   08-NOV-2000; 2000US-0246523P.
PR   08-NOV-2000; 2000US-0246524P.
PR   08-NOV-2000; 2000US-0246525P.
PR   08-NOV-2000; 2000US-0246526P.
PR   08-NOV-2000; 2000US-0246527P.
PR   08-NOV-2000; 2000US-0246528P.
PR   08-NOV-2000; 2000US-0246532P.
PR   08-NOV-2000; 2000US-0246609P.
PR   08-NOV-2000; 2000US-0246610P.
PR   08-NOV-2000; 2000US-0246611P.
PR   08-NOV-2000; 2000US-0246613P.
PR   17-NOV-2000; 2000US-0249207P.
PR   17-NOV-2000; 2000US-0249208P.
PR   17-NOV-2000; 2000US-0249209P.
PR   17-NOV-2000; 2000US-0249210P.
PR   17-NOV-2000; 2000US-0249211P.
PR   17-NOV-2000; 2000US-0249212P.
PR   17-NOV-2000; 2000US-0249213P.
```

```
PR    17-NOV-2000;  2000US-0249214P.
PR    17-NOV-2000;  2000US-0249215P.
PR    17-NOV-2000;  2000US-0249216P.
PR    17-NOV-2000;  2000US-0249217P.
PR    17-NOV-2000;  2000US-0249218P.
PR    17-NOV-2000;  2000US-0249244P.
PR    17-NOV-2000;  2000US-0249245P.
PR    17-NOV-2000;  2000US-0249264P.
PR    17-NOV-2000;  2000US-0249265P.
PR    17-NOV-2000;  2000US-0249297P.
PR    17-NOV-2000;  2000US-0249299P.
PR    17-NOV-2000;  2000US-0249300P.
PR    01-DEC-2000;  2000US-0250160P.
PR    01-DEC-2000;  2000US-0250391P.
PR    05-DEC-2000;  2000US-0251030P.
PR    05-DEC-2000;  2000US-0251988P.
PR    05-DEC-2000;  2000US-0256719P.
PR    06-DEC-2000;  2000US-0251479P.
PR    08-DEC-2000;  2000US-0251856P.
PR    08-DEC-2000;  2000US-0251868P.
PR    08-DEC-2000;  2000US-0251869P.
PR    08-DEC-2000;  2000US-0251989P.
PR    08-DEC-2000;  2000US-0251990P.
PR    11-DEC-2000;  2000US-0254097P.
PR    05-JAN-2001;  2001US-0259678P.
XX
PA    (HUMA-) HUMAN GENOME SCI INC.
XX
PI    Rosen CA,  Barash SC,  Ruben SM;
XX
DR    WPI; 2001-476182/51.
XX
PT    Novel isolated human cytoskeletal element-related polypeptide useful for
PT    diagnosis/treatment of neoplastic disorders, disorders associated with
PT    neural transmission, chromosomal abnormalities, autoimmune disorders.
XX
PS    Claim 1; SEQ ID NO 113; 505pp; English.
XX
CC    Sequences AAS29814-AAS29838 represent genomic DNA molecules, which encode
CC    the cytoskeletal element-related polypeptides of the invention.
CC    Cytoskeletal polypeptides and their associated polynucleotides are useful
CC    in the diagnosis, treatment and prevention of various types of disorders
CC    in e.g. humans, mice, rabbits, goats, horses, cats, dogs, chickens or
CC    sheep. A pathological condition can be determined by determining the
CC    presence or absence of a mutation in a cytoskeletal polynucleotide. The
CC    treatable disorders include autoimmune diseases such as rheumatoid
CC    arthritis, hyperproliferative disorders such as neoplasms of the breast
CC    or liver, cardiovascular disorders such as cardiac arrest,
CC    cerebrovascular disorders such as cerebral ischaemia, nervous system
CC    disorders such as Alzheimer's disease, infections caused by bacteria,
CC    viruses and fungi, ocular disorders such as corneal infection, endocrine
CC    disorders such as premature labour and infertility, gastrointestinal
CC    disorders such as Crohn's disease, renal disorders such as
CC    glomerulonephritis and respiratory disorders such as asthma. The
CC    polypeptides can also be used to aid wound healing, to prevent skin aging
CC    due to sunburn, to maintain organs before transplantation, to regenerate
CC    tissues and in chemotaxis. The polypeptides can also be used as a food
CC    additive or preservative to increase or decrease storage capabilities.
CC    Note:The sequence data for this patent did not form part of the printed
CC    specification, but was obtained in electronic format directly from WIPO
CC    at ftp.wipo.int/pub/published_pct_sequences


Query Match     80.0%;   Score 16.8;  DB 4;  Length 32177;
Best Local Similarity   90.0%;   Pred. No. 2.9e+02;
Matches   18;  Conservative    0;  Mismatches    2;   Indels    0;   Gaps    0;

Qy       1 GGGATTCCTAGAGGGAAGGA 20
           | ||||||||| ||||||||
Db    1139 GAGAATCCTAGAGGGGAAGGA 1120
```

RESULT 10
AAL05374/c

```
ID   AAL05374 standard; DNA; 32177 BP.
XX
AC   AAL05374;
XX
DT   21-NOV-2001  (first entry)
XX
DE   Human reproductive system related antigen DNA SEQ ID NO: 8062.
XX
KW   Human; reproductive system related antigen; reproductive system disorder;
KW   cancer; gene therapy; ds.
XX
OS   Homo sapiens.
XX
PN   WO200155320-A2.
XX
PD   02-AUG-2001.
XX
PF   17-JAN-2001;  2001WO-US001339.
XX
PR   31-JAN-2000;  2000US-0179065P.
PR   04-FEB-2000;  2000US-0180628P.
PR   24-FEB-2000;  2000US-0184664P.
PR   02-MAR-2000;  2000US-0186350P.
PR   16-MAR-2000;  2000US-0189874P.
PR   17-MAR-2000;  2000US-0190076P.
PR   18-APR-2000;  2000US-0198123P.
PR   19-MAY-2000;  2000US-0205515P.
PR   07-JUN-2000;  2000US-0209467P.
PR   28-JUN-2000;  2000US-0214886P.
PR   30-JUN-2000;  2000US-0215135P.
PR   07-JUL-2000;  2000US-0216647P.
PR   07-JUL-2000;  2000US-0216880P.
PR   11-JUL-2000;  2000US-0217487P.
PR   11-JUL-2000;  2000US-0217496P.
PR   14-JUL-2000;  2000US-0218290P.
PR   26-JUL-2000;  2000US-0220963P.
PR   26-JUL-2000;  2000US-0220964P.
PR   14-AUG-2000;  2000US-0224518P.
PR   14-AUG-2000;  2000US-0224519P.
PR   14-AUG-2000;  2000US-0225213P.
PR   14-AUG-2000;  2000US-0225214P.
PR   14-AUG-2000;  2000US-0225266P.
PR   14-AUG-2000;  2000US-0225267P.
PR   14-AUG-2000;  2000US-0225268P.
PR   14-AUG-2000;  2000US-0225270P.
PR   14-AUG-2000;  2000US-0225447P.
PR   14-AUG-2000;  2000US-0225757P.
PR   14-AUG-2000;  2000US-0225758P.
PR   14-AUG-2000;  2000US-0225759P.
PR   18-AUG-2000;  2000US-0226279P.
PR   22-AUG-2000;  2000US-0226681P.
PR   22-AUG-2000;  2000US-0226868P.
PR   22-AUG-2000;  2000US-0271182P.
PR   23-AUG-2000;  2000US-0277009P.
PR   30-AUG-2000;  2000US-0289924P.
PR   01-SEP-2000;  2000US-0292287P.
PR   01-SEP-2000;  2000US-0229343P.
PR   01-SEP-2000;  2000US-0229344P.
PR   01-SEP-2000;  2000US-0229345P.
PR   05-SEP-2000;  2000US-0229509P.
PR   05-SEP-2000;  2000US-0229513P.
PR   06-SEP-2000;  2000US-0230437P.
PR   06-SEP-2000;  2000US-0230438P.
PR   08-SEP-2000;  2000US-0231242P.
PR   08-SEP-2000;  2000US-0231243P.
PR   08-SEP-2000;  2000US-0231244P.
PR   08-SEP-2000;  2000US-0231413P.
PR   08-SEP-2000;  2000US-0231414P.
PR   08-SEP-2000;  2000US-0232080P.
PR   08-SEP-2000;  2000US-0232081P.
PR   12-SEP-2000;  2000US-0231968P.
PR   14-SEP-2000;  2000US-0232397P.
PR   14-SEP-2000;  2000US-0232398P.
```

PR   14-SEP-2000;   2000US-0233399P.
PR   14-SEP-2000;   2000US-0232400P.
PR   14-SEP-2000;   2000US-0232401P.
PR   14-SEP-2000;   2000US-0233063P.
PR   14-SEP-2000;   2000US-0233064P.
PR   14-SEP-2000;   2000US-0233065P.
PR   21-SEP-2000;   2000US-0234223P.
PR   21-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234997P.
PR   25-SEP-2000;   2000US-0234998P.
PR   26-SEP-2000;   2000US-0235484P.
PR   27-SEP-2000;   2000US-0235834P.
PR   27-SEP-2000;   2000US-0235836P.
PR   29-SEP-2000;   2000US-0236327P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239935P.
PR   13-OCT-2000;   2000US-0239937P.
PR   20-OCT-2000;   2000US-0240960P.
PR   20-OCT-2000;   2000US-0241221P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241786P.
PR   20-OCT-2000;   2000US-0241787P.
PR   20-OCT-2000;   2000US-0241808P.
PR   20-OCT-2000;   2000US-0241809P.
PR   20-OCT-2000;   2000US-0241826P.
PR   01-NOV-2000;   2000US-0244617P.
PR   08-NOV-2000;   2000US-0246474P.
PR   08-NOV-2000;   2000US-0246475P.
PR   08-NOV-2000;   2000US-0246476P.
PR   08-NOV-2000;   2000US-0246477P.
PR   08-NOV-2000;   2000US-0246478P.
PR   08-NOV-2000;   2000US-0246523P.
PR   08-NOV-2000;   2000US-0246524P.
PR   08-NOV-2000;   2000US-0246525P.
PR   08-NOV-2000;   2000US-0246526P.
PR   08-NOV-2000;   2000US-0246527P.
PR   08-NOV-2000;   2000US-0246528P.
PR   08-NOV-2000;   2000US-0246532P.
PR   08-NOV-2000;   2000US-0246609P.
PR   08-NOV-2000;   2000US-0246610P.
PR   08-NOV-2000;   2000US-0246611P.
PR   08-NOV-2000;   2000US-0246613P.
PR   08-NOV-2000;   2000US-0249207P.
PR   08-NOV-2000;   2000US-0249208P.
PR   08-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.
PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   17-NOV-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250160P.
PR   01-DEC-2000;   2000US-0250191P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251988P.

PR   05-DEC-2000;   2000US-0256719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,   Barash SC,   Ruben SM;
XX
DR   WPI; 2001-465570/50.
XX
PT   Isolated nucleic acid molecule encoding a reproductive system antigen is
PT   used in preventing, treating or ameliorating a medical condition.
XX
PS   Disclosure; SEQ ID NO 8062; 1297pp + Sequence Listing; English.
XX
CC   The present invention provides the protein and coding sequences of a
CC   number of human reproductive system related antigens. These can be used
CC   in the prevention and treatment of reproductive system disorders,
CC   including cancer. The present sequence is a genomic sequence encoding a
CC   protein of the invention
XX
SQ   Sequence 32177 BP; 8826 A; 6801 C; 7130 G; 9420 T; 0 U; 0 Other;

Query Match   80.0%; Score 16.8; DB 4; Length 32177;
Best Local Similarity   90.0%; Pred. No. 2.9e+02;
Matches   18; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   1 GGGATTCCTAGAGGGAAGGA 20
              | | ||||||||||||||||
Db   1139 GAGAATCCTAGAGGGAAGGA 1120

RESULT 11
ABA18493/c
ID   ABA18493 standard; DNA; 32177 BP.
XX
AC   ABA18493;
XX
DT   23-JAN-2002 (first entry)
XX
DE   Human nervous system related polynucleotide SEQ ID NO 10824.
XX
KW   Human; nootropic; neuroprotective; cytostatic; dermatological; virucide;
KW   immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;
KW   antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;
KW   antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;
KW   antiallergic; antidiabetic; antiulcer; anticonvulsant; antifungal;
KW   antiparasitic; cardiant; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX
OS   Homo sapiens.
XX
PN   WO200159063-A2.
XX
PD   16-AUG-2001.
XX
PF   17-JAN-2001;   2001WO-US001334.
XX
PR   31-JAN-2000;   2000US-0179065P.
PR   04-FEB-2000;   2000US-0180628P.
PR   24-FEB-2000;   2000US-0184664P.
PR   02-MAR-2000;   2000US-0186350P.
PR   16-MAR-2000;   2000US-0189874P.
PR   17-MAR-2000;   2000US-0190076P.
PR   18-APR-2000;   2000US-0198123P.
PR   19-MAY-2000;   2000US-0205515P.
PR   07-JUN-2000;   2000US-0209467P.

```
PR   28-JUN-2000;   2000US-0214886P.          PR   20-OCT-2000;   2000US-0241787P.
PR   30-JUN-2000;   2000US-0215135P.          PR   20-OCT-2000;   2000US-0241808P.
PR   07-JUL-2000;   2000US-0216647P.          PR   20-OCT-2000;   2000US-0241809P.
PR   07-JUL-2000;   2000US-0216880P.          PR   20-OCT-2000;   2000US-0241826P.
PR   11-JUL-2000;   2000US-0217487P.          PR   20-OCT-2000;   2000US-0242221P.
PR   11-JUL-2000;   2000US-0217496P.          PR   01-NOV-2000;   2000US-0244617P.
PR   14-JUL-2000;   2000US-0218290P.          PR   08-NOV-2000;   2000US-0246474P.
PR   26-JUL-2000;   2000US-0220963P.          PR   08-NOV-2000;   2000US-0246475P.
PR   26-JUL-2000;   2000US-0220964P.          PR   08-NOV-2000;   2000US-0246476P.
PR   14-AUG-2000;   2000US-0224518P.          PR   08-NOV-2000;   2000US-0246477P.
PR   14-AUG-2000;   2000US-0224519P.          PR   08-NOV-2000;   2000US-0246478P.
PR   14-AUG-2000;   2000US-0225213P.          PR   08-NOV-2000;   2000US-0246523P.
PR   14-AUG-2000;   2000US-0225214P.          PR   08-NOV-2000;   2000US-0246524P.
PR   14-AUG-2000;   2000US-0225266P.          PR   08-NOV-2000;   2000US-0246525P.
PR   14-AUG-2000;   2000US-0225267P.          PR   08-NOV-2000;   2000US-0246526P.
PR   14-AUG-2000;   2000US-0225268P.          PR   08-NOV-2000;   2000US-0246527P.
PR   14-AUG-2000;   2000US-0225270P.          PR   08-NOV-2000;   2000US-0246528P.
PR   14-AUG-2000;   2000US-0225447P.          PR   08-NOV-2000;   2000US-0246532P.
PR   14-AUG-2000;   2000US-0225757P.          PR   08-NOV-2000;   2000US-0246609P.
PR   14-AUG-2000;   2000US-0225758P.          PR   08-NOV-2000;   2000US-0246610P.
PR   14-AUG-2000;   2000US-0225759P.          PR   08-NOV-2000;   2000US-0246611P.
PR   18-AUG-2000;   2000US-0226279P.          PR   08-NOV-2000;   2000US-0246613P.
PR   22-AUG-2000;   2000US-0226681P.          PR   17-NOV-2000;   2000US-0249207P.
PR   22-AUG-2000;   2000US-0226868P.          PR   17-NOV-2000;   2000US-0249208P.
PR   22-AUG-2000;   2000US-0227182P.          PR   17-NOV-2000;   2000US-0249209P.
PR   23-AUG-2000;   2000US-0227009P.          PR   17-NOV-2000;   2000US-0249210P.
PR   30-AUG-2000;   2000US-0228924P.          PR   17-NOV-2000;   2000US-0249211P.
PR   01-SEP-2000;   2000US-0229287P.          PR   17-NOV-2000;   2000US-0249212P.
PR   01-SEP-2000;   2000US-0229343P.          PR   17-NOV-2000;   2000US-0249213P.
PR   01-SEP-2000;   2000US-0229344P.          PR   17-NOV-2000;   2000US-0249214P.
PR   01-SEP-2000;   2000US-0229345P.          PR   17-NOV-2000;   2000US-0249215P.
PR   05-SEP-2000;   2000US-0229509P.          PR   17-NOV-2000;   2000US-0249216P.
PR   05-SEP-2000;   2000US-0229513P.          PR   17-NOV-2000;   2000US-0249217P.
PR   06-SEP-2000;   2000US-0230437P.          PR   17-NOV-2000;   2000US-0249218P.
PR   06-SEP-2000;   2000US-0230438P.          PR   17-NOV-2000;   2000US-0249244P.
PR   08-SEP-2000;   2000US-0231242P.          PR   17-NOV-2000;   2000US-0249245P.
PR   08-SEP-2000;   2000US-0231243P.          PR   17-NOV-2000;   2000US-0249264P.
PR   08-SEP-2000;   2000US-0231244P.          PR   17-NOV-2000;   2000US-0249265P.
PR   08-SEP-2000;   2000US-0231413P.          PR   17-NOV-2000;   2000US-0249297P.
PR   08-SEP-2000;   2000US-0231414P.          PR   17-NOV-2000;   2000US-0249299P.
PR   08-SEP-2000;   2000US-0232080P.          PR   17-NOV-2000;   2000US-0249300P.
PR   08-SEP-2000;   2000US-0232081P.          PR   01-DEC-2000;   2000US-0250391P.
PR   12-SEP-2000;   2000US-0231968P.          PR   01-DEC-2000;   2000US-0251160P.
PR   14-SEP-2000;   2000US-0232397P.          PR   05-DEC-2000;   2000US-0251030P.
PR   14-SEP-2000;   2000US-0232398P.          PR   05-DEC-2000;   2000US-0251988P.
PR   14-SEP-2000;   2000US-0232399P.          PR   05-DEC-2000;   2000US-0267519P.
PR   14-SEP-2000;   2000US-0232400P.          PR   05-DEC-2000;   2000US-0567719P.
PR   14-SEP-2000;   2000US-0232401P.          PR   06-DEC-2000;   2000US-0251479P.
PR   14-SEP-2000;   2000US-0233063P.          PR   08-DEC-2000;   2000US-0251856P.
PR   14-SEP-2000;   2000US-0233064P.          PR   08-DEC-2000;   2000US-0251868P.
PR   14-SEP-2000;   2000US-0233065P.          PR   08-DEC-2000;   2000US-0251869P.
PR   21-SEP-2000;   2000US-0234223P.          PR   08-DEC-2000;   2000US-0251989P.
PR   21-SEP-2000;   2000US-0234274P.          PR   08-DEC-2000;   2000US-0251990P.
PR   25-SEP-2000;   2000US-0234997P.          PR   11-DEC-2000;   2000US-0540097P.
PR   25-SEP-2000;   2000US-0234998P.          PR   05-JAN-2001;   2001US-0259678P.
PR   26-SEP-2000;   2000US-0235484P.          XX
PR   27-SEP-2000;   2000US-0235834P.          PA   (HUMA-) HUMAN GENOME SCI INC.
PR   27-SEP-2000;   2000US-0235836P.          XX
PR   29-SEP-2000;   2000US-0236327P.          PI   Rosen CA,   Barash SC,   Ruben SM;
PR   29-SEP-2000;   2000US-0236367P.          XX
PR   29-SEP-2000;   2000US-0236368P.          DR   WPI; 2001-541565/60.
PR   29-SEP-2000;   2000US-0236369P.          XX
PR   29-SEP-2000;   2000US-0236370P.          PT   Nucleic acids encoding 3224 human nervous system antigen polypeptides,
PR   02-OCT-2000;   2000US-0236802P.          PT   useful for preventing, diagnosing and/or treating nervous system cancers
PR   02-OCT-2000;   2000US-0237037P.          PT   and metastases.
PR   02-OCT-2000;   2000US-0237038P.          XX
PR   02-OCT-2000;   2000US-0237039P.          PS   Disclosure; SEQ ID NO 10824; 1701pp + Sequence Listing; English.
PR   02-OCT-2000;   2000US-0237040P.          XX
PR   13-OCT-2000;   2000US-0239335P.          CC   The invention relates to novel genes (ABA11004-ABA21534) and proteins
PR   13-OCT-2000;   2000US-0239337P.          CC   (ABB14678-ABB18001) useful for preventing, treating or ameliorating
PR   20-OCT-2000;   2000US-0240960P.          CC   medical conditions e.g. by protein or gene therapy. The genes are
PR   20-OCT-2000;   2000US-0241785P.          CC   isolated from a range of human tissues disclosed in the specification.
PR   20-OCT-2000;   2000US-0241786P.          CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful in
                                              CC   the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and
```

```
CC     ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,
CC     breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune
CC     disorders e.g. Addison's disease, allergies, autoimmune haemolytic
CC     anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,
CC     multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)
CC     cardiovascular disorders such as myocardial ischaemias; (d) wound healing
CC     ; (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)
CC     infectious diseases such as viral, bacterial, fungal and parasitic
CC     infections. Note: The sequence data for this patent did not form part of
CC     the printed specification, but was obtained in electronic format directly
CC     from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 32177 BP; 8826 A; 6801 C; 7130 G; 9420 T; 0 U; 0 Other;

Query Match             18;
Best Local Similarity   90.0%;    Score 16.8;  DB 5; Length 32177;
                                  Pred. No. 2.9e+02;
Matches  18;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy       1 GGATTCCTAGAGGGAAGGAGGA 20
           ||| |||||||||||||||||
Db    1139 GAGAATCCTAGAGGGAAGGA 1120

RESULT 12
AAX91990_05/c
Continuation (6 of 13) of AAX91990 from base 500001 (Nucleotide sequence of the complete
WP Sequence split into 13 fragments LOCUS AAX91990 Accession Aax91990
WP Fragment Name      Begin        End
WP AAX91990_00             1     110000
WP AAX91990_01        100001     210000
WP AAX91990_02        200001     310000
WP AAX91990_03        300001     410000
WP AAX91990_04        400001     510000
WP AAX91990_05        500001     610000
WP AAX91990_06        600001     710000
WP AAX91990_07        700001     810000
WP AAX91990_08        800001     910000
WP AAX91990_09        900001    1010000
WP AAX91990_10       1000001    1110000
WP AAX91990_11       1100001    1210000
WP AAX91990_12       1200001    1230025

Query Match
Best Local Similarity   90.0%;    Score 16.8;  DB 2; Length 110000;
                                  Pred. No. 3.1e+02;
Matches  18;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy       2 GGATTCCTAGAGGGAAGGAGGAG 21
           |||||||||||||| ||||
Db   46344 GGATTCCTACAGGGAGGGAG 46325

RESULT 13
AAC81914
ID   AAC81914 standard; DNA; 273254 BP.
XX
AC   AAC81914;
XX
DT   15-SEP-2003  (revised)
DT   27-FEB-2001  (first entry)
XX
DE   Chlamydia pneumoniae genome DNA.
XX
KW   Genome; diagnosis; vaccine; ds.
XX
OS   Chlamydophila pneumoniae.
XX
PN   WO200027994-A2.
XX
PD   18-MAY-2000.
XX
PF   12-NOV-1999;  99WO-US026923.
XX
PR   12-NOV-1998;  98US-0108279P.
```

```
PR   08-APR-1999;  99US-0128606P.
XX
PA   (REGC ) UNIV CALIFORNIA.
XX
PI   Stephens R, Mitchell W, Kalman S, Davis R;
XX
DR   WPI; 2000-376516/32.
XX
PT   Isolated nucleic acid for use in diagnostic and analytical methods
PT   encodes genomic sequence of Chlamydia pneumoniae.
XX
PS   Claim 2; Page 128-320; 320pp; English.
XX
CC   This invention describes a novel nucleic acid (N1) encoding a Chlamydia
CC   pneumoniae protein (P1), given in the specification. The isolated nucleic
CC   acid is useful for diagnostic and analytical methods, such as,
CC   hybridization-based assays or amplification-based assays. The protein may
CC   be used for diagnostic purposes, for their enzymatic or structural
CC   activity, or as a vaccine. The invention also describes (1) a probe
CC   comprising a hybridizing fragment of N1; (2) an isolated nucleic acid
CC   (N2) that hybridizes under stringent conditions to N1; (3) an expression
CC   cassette comprising N1 under the transcriptional regulation of a
CC   transcriptional initiation region functional in an expression host, and a
CC   transcriptional termination region; (4) a cell comprising an expression
CC   cassette of (3) as part of an extrachromosomal element or integrated into
CC   the genome of a host cell as a result of induction of the expression
CC   cassette into the host cell, and the cellular progeny of the host cell;
CC   (5) a method for producing a P1 comprising growing a cell of (4) where
CC   the protein is expressed and isolating the protein free of other proteins
CC   ; (6) a purified polypeptide composition comprising at least 50 weight %
CC   of P1; and (7) a monoclonal antibody binding specifically to the peptide
CC   of (6). (Updated on 15-SEP-2003 to standardise OS field)
XX
SQ   Sequence 273254 BP; 76423 A; 51054 C; 61965 G; 83812 T; 0 U; 0 Other;

Query Match
Best Local Similarity   90.0%;    Score 16.8;  DB 3; Length 273254;
                                  Pred. No. 3.2e+02;
Matches  18;  Conservative  0;  Mismatches  2;  Indels  0;  Gaps  0;

Qy       2 GGATTCCTAGAGGGAAGGAG 21
           |||||||||| ||||
Db   70053 GGATTCCTACAGGGAGGGAG 70072

RESULT 14
AAS43104
ID   AAS43104 standard; DNA; 325791 BP.
XX
AC   AAS43104;
XX
DT   18-DEC-2001  (first entry)
XX
DE   Human Oestrogen receptor beta gene.
XX
KW   Human; Oestrogen receptor beta; ERbeta; ds; SNP; chromosome 6q.25.1;
KW   single nucleotide polymorphism; cardiovascular disease;
KW   autoimmune disease; systemic lupus erythematosus; arthritis; rheumatism;
KW   osteoarthritis; osteoporosis; breast cancer; endometrial cancer.
XX
OS   Homo sapiens.
XX
PN   WO200162793-A2.
XX
PD   30-AUG-2001.
XX
PF   20-FEB-2001;  2001WO-US005360.
XX
PR   22-FEB-2000;  2000US-0183755P.
PR   24-JAN-2001;  2001US-00768185.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Kalush F, Cassel MJ, Hwang SS, Winn-Deen ES;
```

```
XX
DR   WPI; 2001-582041/65.
DR   P-PSDB; AAU27322.
XX
PT   Estrogen receptor gene and protein polymorphisms useful for diagnosis of
PT   individuals at risk of developing bone disorders.
XX
PS   Example 2; Fig 1; 245pp; English.
XX
CC   The invention relates to a novel isolated peptide comprising or
CC   consisting of an amino acid sequence selected from an amino acid sequence
CC   of a variant oestrogen receptor protein (e.g. ERbeta), or a fragment of
CC   10 amino acids), antibodies against them, nucleic acids encoding them
CC   (including vectors for transforming cells). The gene for human ERbeta is
CC   located on chromosome 6q.25.1. The variants are encoded by single
CC   nucleotide polymorphisms (SNP). The variant peptides and proteins can be
CC   used in assays to determine the biological activity of the protein, to
CC   raise antibodies, as a reagent in assays designed to quantitatively
CC   determine levels of the protein in biological fluids, to identify
CC   compounds that modulate receptor activity and to screen compounds for the
CC   ability to stimulate or inhibit interaction between the receptor protein
CC   and a target molecule that normally interacts with the receptor protein
CC   e.g. oestrogen. The antibody can be used to isolate the protein, to
CC   assess expression in disease states e.g. cardiovascular disease and
CC   autoimmune disease (e.g. systemic lupus erythematosus, arthritis,
CC   rheumatism and osteoarthritis), osteoporosis, breast cancer and
CC   endometerial cancer. In addition the antibodies can be used in
CC   pharmacogenomic analysis and inhibiting protein function, e.g. blocking
CC   the binding of the oestrogen receptor protein to a binding partner such
CC   as a ligand. The nucleic acids encoding the proteins can be used as
CC   probes, primers, chemical intermediates and in biological assays. The
CC   present sequence is the human ERbeta gene
XX
SQ   Sequence 325791 BP; 94098 A; 68292 C; 67970 G; 95431 T; 0 U; 0 Other;

Query Match       80.0%;  Score 16.8;  DB 4;  Length 325791;
Best Local Similarity   90.0%;  Pred. No. 3.3e+02;
Matches   18;  Conservative   0;  Mismatches   2;  Indels   0;  Gaps   0;

Qy      2 GGATTCCTAGAGGGAAGGAG 21
          |||||||||||||| ||||
Db 263062 GGATTCCTAGAGGCCAGGAG 263081


RESULT 15
AAC24739/c
ID   AAC24739 standard; cDNA; 303 BP.
XX
AC   AAC24739;
XX
DT   06-OCT-2000 (first entry)
XX
DE   Human secreted protein 5' EST, SEQ ID NO: 28814.
XX
KW   Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;
KW   gene therapy; chromosome mapping; ss.
XX
OS   Homo sapiens.
XX
PN   EP1033401-A2.
XX
PD   06-SEP-2000.
XX
PF   21-FEB-2000; 2000EP-00200610.
XX
PR   26-FEB-1999;  99US-0122487P.
XX
PA   (GEST ) GENSET.
XX
PI   Dumas Milne Edwards J, Duclert A, Giordano J;
XX
DR   WPI; 2000-500381/45.
XX
PT   New nucleic acid that is a 5' expressed sequence tag (5' EST) for
PT   obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for
PT   diagnostic, forensic, gene therapy and chromosome mapping procedures.
XX
PS   Claim 1; SEQ ID NO 28814; 71pp + Sequence Listing; English.
XX
CC   The present sequence is one of a large number of 5' ESTs derived from
CC   mRNAs encoding secreted proteins. No ORF has yet been conclusively
CC   identified within the present sequence. The 5' ESTs were prepared from
CC   total human RNAs or polyA+ RNAs derived from 30 different tissues. EST
CC   sequences usually correspond mainly to the 3' untranslated region (UTR)
CC   of the mRNA because they are often obtained from oligo-dT primed cDNA
CC   libraries. Such ESTs are not well suited for isolating cDNA sequences
CC   derived from the 5' ends of mRNAs and even in those cases where longer
CC   cDNA sequences have been obtained, the full 5' UTR is rarely included. 5'
CC   ESTs are derived from mRNAs with intact 5' ends and can therefore be used
CC   to obtain full length cDNAs and genomic DNAs. 5' ESTs are also used in
CC   diagnostic, forensic, gene therapy and chromosome mapping procedures.
CC   They are used to obtain upstream regulatory sequences and to design
CC   expression and secretion vectors
XX
SQ   Sequence 303 BP; 86 A; 66 C; 44 G; 99 T; 0 U; 8 Other;

Query Match       78.1%;  Score 16.4;  DB 3;  Length 303;
Best Local Similarity   94.4%;  Pred. No. 3.4e+02;
Matches   17;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy      4 ATTCCTAGAGGGAAGGAG 21
          ||| ||||||||||||||
Db    177 ATTGCTAGAGGGAAGGAG 160

Search completed: March 25, 2004, 10:25:24
Job time : 25.7931 secs
```

OM nucleic - nucleic search, using sw model

```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004  Compugen Ltd.
```

Run on: March 25, 2004, 09:55:14 ; Search time 4.13061 seconds
                         (without alignments)
                         2821.370 Million cell updates/sec

Title: US-09-963-285-1_COPY_403_423
Perfect score: 21
Sequence: 1 gggattcctagaggaggag 21

Scoring table: IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :   Issued Patents NA:*
             1: /cgn2_6/ptcdata/2/ina/5A_COMB.seq:*
             2: /cgn2_6/ptcdata/2/ina/5B_COMB.seq:*
             3: /cgn2_6/ptcdata/2/ina/6A_COMB.seq:*
             4: /cgn2_6/ptcdata/2/ina/6B_COMB.seq:*
             5: /cgn2_6/ptcdata/2/ina/PCTUS_COMB.seq:*
             6: /cgn2_6/ptcdata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	16.8	80.0	1230025	4	US-09-198-452A-1	Sequence 1, Appli
2	16.4	78.1	202001	4	US-09-734-674-3	Sequence 3, Appli
3	16.2	77.1	900	4	US-09-393-634-44	Sequence 44, Appli
4	16.2	77.1	38653	4	US-09-922-445-1	Sequence 1, Appli
5	15.8	75.2	287	4	US-09-313-294A-1436	Sequence 1436, Ap
6	15.8	75.2	955	4	US-09-641-638-13	Sequence 13, Appl
7	15.8	75.2	955	4	US-09-641-638-14	Sequence 14, Appl
8	15.8	75.2	43069	2	US-09-292-542A-1	Sequence 1, Appl
9	15.4	73.3	3502	2	US-08-724-394A-16	Sequence 16, Appl
10	15.2	72.4	288	4	US-09-621-976-12965	Sequence 12965, A
11	15.2	72.4	351	4	US-09-621-976-10679	Sequence 10679, A
12	15.2	72.4	463	4	US-09-621-976-2535	Sequence 2535, Ap
13	15.2	72.4	554	4	US-09-621-976-18505	Sequence 18505, A
14	15.2	72.4	642	4	US-09-252-991A-13569	Sequence 13569, A
15	15.2	72.4	810	4	US-09-252-991A-12999	Sequence 12999, A
16	15.2	72.4	873	4	US-09-252-991A-12472	Sequence 12472, A
17	15.2	72.4	893	3	US-09-313-300-5	Sequence 5, Appli
18	15.2	72.4	1001	4	US-09-671-317-149	Sequence 149, App
19	15.2	72.4	1686	4	US-09-833-381-2019	Sequence 2019, Ap
20	15.2	72.4	1707	4	US-09-596-248D-22	Sequence 22, Appl
21	15.2	72.4	1766	4	US-09-517-467B-18	Sequence 18, Appl
22	15.2	72.4	1787	4	US-08-923-454A-3	Sequence 3, Appli
23	15.2	72.4	2526	1	US-07-912-952-1	Sequence 1, Appl
24	15.2	72.4	2551	3	US-08-923-454A-26	Sequence 26, Appl
25	15.2	72.4	3000	4	US-09-705-267A-18	Sequence 18, Appl
26	15.2	72.4	3033	4	US-09-833-381-1810	Sequence 1810, Ap
27	15.2	72.4	3129	4	US-09-252-991A-13873	Sequence 13873, A
28	15.2	72.4	4131	4	US-09-252-991A-13773	Sequence 13773, A
29	15.2	72.4	8211	4	US-09-252-991A-13656	Sequence 13656, A
30	15.2	72.4	12425	4	US-09-616-289-50	Sequence 50, Appl
31	15.2	72.4	15652	4	US-09-422-936-60	Sequence 60, Appl
32	15.2	72.4	18000	4	US-09-657-346A-17	Sequence 17, App
33	15.2	72.4	29604	3	US-08-781-891-207	Sequence 207, App
34	15.2	72.4	29604	4	US-09-618-166-207	Sequence 207, App
35	15.2	72.4	87350	3	US-08-781-891-79	Sequence 79, Appl
36	15.2	72.4	87350	4	US-09-618-166-79	Sequence 79, Appl
37	15.2	72.4	87543	4	US-09-791-211-3	Sequence 3, Appli
38	15.2	72.4	161652	4	US-09-497-855A-40	Sequence 40, Appl
39	15.2	72.4	169998	4	US-09-677-610B-24	Sequence 24, Appl
40	15.2	72.4	197496	4	US-09-877-177A-10	Sequence 10, Appl
41	14.8	70.5	512	4	US-09-621-976-3594	Sequence 3594, Ap
42	14.8	70.5	1376	4	US-09-443-184-44	Sequence 44, Appl
43	14.8	70.5	1390	4	US-09-205-258-124	Sequence 124, App
44	14.8	70.5	2935	4	US-09-976-594-365	Sequence 365, App
45	14.8	70.5	2935	4	US-09-976-594-365	Sequence 365, App

ALIGNMENTS

```
RESULT 1
US-09-198-452A-1/c
; Sequence 1, Application US/09198452A
; Patent No. 6559294
; GENERAL INFORMATION:
; APPLICANT: Griffais, R.
; TITLE OF INVENTION: Chlamydia pneumoniae genomic sequence and polypeptides, fragments
; TITLE OF INVENTION: thereof and uses thereof, in particular for the diagnosis, preve
; TITLE OF INVENTION: and treatment of infection
; FILE REFERENCE: 9710-003-999
; CURRENT APPLICATION NUMBER: US/09/198,452A
; CURRENT FILING DATE: 1998-11-24
; NUMBER OF SEQ ID NOS: 6849
SEQ ID NO 1
; LENGTH: 1230025
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(15000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (15001)..(30000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (30001)..(45000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (45001)..(60000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (60001)..(75000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (75001)..(90000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (90001)..(105000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (105001)..(120000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (120001)..(135000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (135001)..(150000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (150001)..(165000)
```

```
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (165001)..(180000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (180001)..(195000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (195001)..(210000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (210001)..(225000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (225001)..(240000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (240001)..(255000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (255001)..(270000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (270001)..(285000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (285001)..(300000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (300001)..(315000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (315001)..(330000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (330001)..(345000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (345001)..(360000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (360001)..(375000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (375001)..(390000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (390001)..(405000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (405001)..(420000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (420001)..(435000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (435001)..(450000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (450001)..(465000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (465001)..(480000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (480001)..(495000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (495001)..(510000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (510001)..(525000)
OTHER INFORMATION: n=a or c or g or t

NAME/KEY: misc_feature
LOCATION: (525001)..(540000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (540001)..(555000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (555001)..(570000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (570001)..(585000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (585001)..(600000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (600001)..(615000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (615001)..(630000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (630001)..(645000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (645001)..(660000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (660001)..(675000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (675001)..(690000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (690001)..(705000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (705001)..(720000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (720001)..(735000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (735001)..(750000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (750001)..(765000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (765001)..(780000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (780001)..(795000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (795001)..(810000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (810001)..(825000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (825001)..(840000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (840001)..(855000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (855001)..(870000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
LOCATION: (870001)..(885000)
OTHER INFORMATION: n=a or c or g or t
NAME/KEY: misc_feature
```

; LOCATION: (885001)..(900000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature
; LOCATION: (900001)..(915000)
; OTHER INFORMATION: n=a or c or g or t
; NAME/KEY: misc_feature

Query Match   80.0%;   Score 16.8; DB 4; Length 1230025;
Best Local Similarity   90.0%;   Pred. No. 56;
Matches   18; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy   2 GGATTCCTAGAGGGAGGAG 21
            ||||||||| ||||||||
Db   546344 GGATTCCTACAGGGAGGAG 546325

RESULT 2
US-09-734-674-3/c
; Sequence 3, Application US/09734674
; Patent No. 6498022
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001018
; CURRENT APPLICATION NUMBER: US/09/734,674
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 202001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(202001)
; OTHER INFORMATION: n = A,T,C or G
US-09-734-674-3

Query Match   78.1%;   Score 16.4; DB 4; Length 202001;
Best Local Similarity   94.4%;   Pred. No. 72;
Matches   17; Conservative 0; Mismatches 1; Indels 0; Gaps .0;

Qy   4 ATTCCTAGAGGGAAGGAG 21
            ||| |||||||||||||
Db   124677 ATTTCTAGAGGGAAGGAG 124660

RESULT 3
US-09-393-634-44;
; Sequence 44, Application US/09393634
; Patent No. 6558910
; GENERAL INFORMATION:
; APPLICANT: Zuker, Charles S.
; APPLICANT: Adler, Jon Elliot
; APPLICANT: Ryba, Nick
; APPLICANT: Mueller, Ken
; APPLICANT: Hoon, Mark
; APPLICANT: The Regents of the University of California
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented by the Secretary of the
; APPLICANT: Department of Health and Human Services
; TITLE OF INVENTION: SF, a No. 6558910e1 Family of Taste Receptors
; FILE REFERENCE: 02307E-098000US
; CURRENT APPLICATION NUMBER: US/09/393,634
; CURRENT FILING DATE: 1999-09-10
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 44
; LENGTH: 900
; TYPE: DNA
; ORGANISM: Homo sapiens

; FEATURE:
; OTHER INFORMATION: human GR05
US-09-393-634-44

Query Match   77.1%;   Score 16.2; DB 4; Length 900;
Best Local Similarity   85.7%;   Pred. No. 41;
Matches   18; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy   1 GGGATTCCTAGAGGGAAGGAG 21
            ||||||||||||||||| |||
Db   823 GGGATTCCTAGGGTGAAGCAG 843

RESULT 4
US-09-922-445-1/c
; Sequence 1, Application US/09922445
; Patent No. 6528268
; GENERAL INFORMATION:
; APPLICANT: Andersson, Maria K.
; APPLICANT: Berglund, Lars G. T.
; APPLICANT: Reneland, Rikard H.
; APPLICANT: Adam, Gail I. R.
; TITLE OF INVENTION: REAGENTS AND METHODS FOR DETECTION OF HEART FAILURE
; FILE REFERENCE: GG126US
; CURRENT APPLICATION NUMBER: US/09/922,445
; CURRENT FILING DATE: 2001-08-03
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 38653
; TYPE: DNA
; ORGANISM: homo sapiens
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: (1)..(26156)
; OTHER INFORMATION:
; NAME/KEY: misc_feature
; LOCATION: (24801)..(24801)
; OTHER INFORMATION: nucleotide 24801 is a single nucleotide polymorphism which can be
; OTHER INFORMATION: A or G
; NAME/KEY: misc_feature
; LOCATION: (24941)..(24941)
; OTHER INFORMATION: nucleotide 24941 is a single nucleotide polymorphism which can be
; OTHER INFORMATION: T or C
; NAME/KEY: exon
; LOCATION: (26157)..(26252)
; OTHER INFORMATION:
; NAME/KEY: Intron
; LOCATION: (26253)..(26401)
; OTHER INFORMATION:
; NAME/KEY: exon
; LOCATION: (26402)..(26543)
; OTHER INFORMATION:
; NAME/KEY: Intron
; LOCATION: (26544)..(27024)
; OTHER INFORMATION:
; NAME/KEY: exon
; LOCATION: (27025)..(27178)
; OTHER INFORMATION:
; NAME/KEY: Intron
; LOCATION: (27179)..(30519)
; OTHER INFORMATION:
; NAME/KEY: misc_feature
; LOCATION: (27645)..(27645)
; OTHER INFORMATION: nucleotide 27645 is a single nucleotide polymorphism which can be
; OTHER INFORMATION: C or G
; NAME/KEY: exon
; LOCATION: (30520)..(30681)
; OTHER INFORMATION:
; NAME/KEY: Intron
; LOCATION: (30682)..(30894)
; OTHER INFORMATION:
; NAME/KEY: exon

```
;   LOCATION: (30895)..(31027)
;   OTHER INFORMATION:
;   NAME/KEY: Intron
;   LOCATION: (31028)..(31747)
;   OTHER INFORMATION:
;   NAME/KEY: exon
;   LOCATION: (31748)..(31841)
;   OTHER INFORMATION:
;   NAME/KEY: Intron
;   LOCATION: (31842)..(32400)
;   OTHER INFORMATION:
;   NAME/KEY: misc_feature
;   LOCATION: (32163)..(32163)
;   OTHER INFORMATION: nucleotide 32163 is a single nucleotide polymorphism which can be
;   OTHER INFORMATION: A or C
;   NAME/KEY: exon
;   LOCATION: (32401)..(32528)
;   OTHER INFORMATION:
;   NAME/KEY: Intron
;   LOCATION: (32529)..(33414)
;   OTHER INFORMATION:
;   NAME/KEY: misc_feature
;   LOCATION: (32614)..(32614)
;   OTHER INFORMATION: nucleotide 32614 is a single nucleotide polymorphism which can be
;   OTHER INFORMATION: A or G
;   NAME/KEY: exon
;   LOCATION: (33415)..(33597)
;   OTHER INFORMATION:
;   NAME/KEY: Intron
;   LOCATION: (33598)..(34314)
;   OTHER INFORMATION:
;   NAME/KEY: exon
;   LOCATION: (34315)..(34588)
;   OTHER INFORMATION:
;   NAME/KEY: Intron
;   LOCATION: (34589)..(36404)
;   OTHER INFORMATION:
;   NAME/KEY: exon
;   LOCATION: (36405)..(36523)
;   OTHER INFORMATION:
;   NAME/KEY: Intron
;   LOCATION: (36524)..(38341)
;   OTHER INFORMATION:
;   NAME/KEY: exon
;   LOCATION: (38342)..(38653)
;   OTHER INFORMATION:
;   PUBLICATION INFORMATION:
;   DATABASE ACCESSION NUMBER: Genbank/AC004923
;   DATABASE ENTRY DATE: 1999-12-21
;   RELEVANT RESIDUES: (1)..(38653)
US-09-922-445-1

Query Match             77.1%;  Score 16.2;  DB 4;  Length 38653;
Best Local Similarity   85.7%;  Pred. No. 72;
Matches  18;  Conservative   0;  Mismatches   3;  Indels   0;  Gaps   0;

Qy      1 GGGATTCCTAGAGGGAAGGAG 21
            ||||  ||  ||||||||||
Db  15818 GGGACTCGGAGAGGGAAGGAG 15798

RESULT 5
US-09-313-294A-1436
; Sequence 1436, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;   APPLICANT: Lalgudi, Raghunath V.
;   APPLICANT: Ito, Laura Y.
;   APPLICANT: Sherman, Bradley K.
;   TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;   FILE REFERENCE: PL-0017 US
;   CURRENT APPLICATION NUMBER: US/09/313,294A
;   CURRENT FILING DATE: 1999-05-14
```

```
;   NUMBER OF SEQ ID NOS: 7600
;   SOFTWARE: PERL Program
; SEQ ID NO 1436
;   LENGTH: 287
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. 6476212 700550945H1
;   NAME/KEY: unsure
;   LOCATION: 109-110
;   OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-1436

Query Match             75.2%;  Score 15.8;  DB 4;  Length 287;
Best Local Similarity   89.5%;  Pred. No. 54;
Matches  17;  Conservative   0;  Mismatches   2;  Indels   0;  Gaps   0;

Qy      3 GATTCCTAGAGGGAAGGAG 21
            || ||||||| ||||||
Db    215 GAATCCTAGATGGAAGGAG 233

RESULT 6
US-09-641-638-13/c
; Sequence 13, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
;   APPLICANT: Blumenfeld, Marta
;   APPLICANT: Bougueleret, Lydie
;   APPLICANT: Chumakov, Ilya
;   APPLICANT: Cohen, Annick
;   TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
;   TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
;   FILE REFERENCE: GENSET.051CP1
;   CURRENT APPLICATION NUMBER: US/09/641,638
;   CURRENT FILING DATE: 2000-08-16
;   PRIOR APPLICATION NUMBER: US 09/502,330
;   PRIOR FILING DATE: 2000-02-11
;   PRIOR APPLICATION NUMBER: US 60/133,200
;   PRIOR FILING DATE: 1999-05-07
;   PRIOR APPLICATION NUMBER: US 09/275,267
;   PRIOR FILING DATE: 1999-03-23
;   PRIOR APPLICATION NUMBER: US 60/119,917
;   PRIOR FILING DATE: 1999-02-12
;   NUMBER OF SEQ ID NOS: 1304
;   SOFTWARE: Patent.pm
; SEQ ID NO 13
;   LENGTH: 955
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 478
;   OTHER INFORMATION: 10-33-175 : polymorphic base C or T
;   NAME/KEY: misc_binding
;   LOCATION: 459..477
;   OTHER INFORMATION: 10-33-175.mis1
;   NAME/KEY: misc_binding
;   LOCATION: 479..498
;   OTHER INFORMATION: 10-33-175.mis2, potential complement
;   NAME/KEY: primer_bind
;   LOCATION: 304..322
;   OTHER INFORMATION: upstream amplification primer
;   NAME/KEY: primer_bind
;   LOCATION: 705..723
;   OTHER INFORMATION: downstream amplification primer, complement
;   NAME/KEY: misc_binding
;   LOCATION: 466..490
;   OTHER INFORMATION: 10-33-175 potential probe
US-09-641-638-13

Query Match             75.2%;  Score 15.8;  DB 4;  Length 955;
```

Best Local Similarity 89.5%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy   2 GGATTCCTAGAGGAAGGA GA 20
           |||||| ||||||||| ||
Db   55 GGATTCCTTGAGGGAATGA 37

RESULT 7
US-09-641-638-14/c
; Sequence 14, Application US/09641638
; Patent No. 6432648
; GENERAL INFORMATION:
;   APPLICANT: Blumenfeld, Marta
;   APPLICANT: Bougueleret, Lydie
;   APPLICANT: Chumakov, Ilya
;   APPLICANT: Cohen, Annick
;   TITLE OF INVENTION: BIALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
;   TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
;   FILE REFERENCE: GENSET.051CP1
;   CURRENT APPLICATION NUMBER: US/09/641,638
;   CURRENT FILING DATE: 2000-08-16
;   PRIOR APPLICATION NUMBER: US 09/502,330
;   PRIOR FILING DATE: 2000-02-11
;   PRIOR APPLICATION NUMBER: US 60/133,200
;   PRIOR FILING DATE: 1999-05-07
;   PRIOR APPLICATION NUMBER: US 09/275,267
;   PRIOR FILING DATE: 1999-03-23
;   PRIOR APPLICATION NUMBER: US 60/119,917
;   PRIOR FILING DATE: 1999-02-12
;   NUMBER OF SEQ ID NOS: 1304
;   SOFTWARE: Patent.pm
; SEQ ID NO 14
;   LENGTH: 955
;   TYPE: DNA
;   ORGANISM: Homo Sapiens
;   FEATURE:
;   NAME/KEY: allele
;   LOCATION: 478
;   OTHER INFORMATION: 10-33-211 : polymorphic base C or T
;   NAME/KEY: misc_binding
;   LOCATION: 458..477
;   OTHER INFORMATION: 10-33-211.mis1, potential
;   NAME/KEY: misc_binding
;   LOCATION: 479..498
;   OTHER INFORMATION: 10-33-211.mis2, potential complement
;   NAME/KEY: primer_bind
;   LOCATION: 268..286
;   OTHER INFORMATION: upstream amplification primer
;   NAME/KEY: primer_bind
;   LOCATION: 669..687
;   OTHER INFORMATION: downstream amplification primer, complement
;   NAME/KEY: misc_binding
;   LOCATION: 466..490
;   OTHER INFORMATION: 10-33-211 potential probe
US-09-641-638-14

Query Match 75.2%; Score 15.8; DB 4; Length 955;
Best Local Similarity 89.5%; Pred. No. 65;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy   2 GGATTCCTAGAGGAAGGA GA 20
           |||||| ||||||||| ||
Db   19 GGATTCCTTGAGGGAATGA 1

RESULT 8
US-09-292-542A-1/c
; Sequence 1, Application US/09292542A
; Patent No. 6551279
; GENERAL INFORMATION:
;   APPLICANT: Blumenfeld, Marta
;   APPLICANT: Chumakov, Ilya

;   APPLICANT: Bougueleret, Lydie
;   TITLE OF INVENTION: Genomic Sequence Of The 5-Lipoxygenase-Activating Protein (FLAP)
;   Patent No. 6551279
;   TITLE OF INVENTION: Polymorphic Markers Thereof And Methods For Detection Of Asthma.
;   FILE REFERENCE: GENSET.026A
;   CURRENT FILING DATE: 1999-04-15
;   CURRENT APPLICATION NUMBER: US/09/292,542A
;   PRIOR APPLICATION NUMBER: US 60/081893
;   PRIOR FILING DATE: 1998-04-15
;   PRIOR APPLICATION NUMBER: US 60/091314
;   PRIOR FILING DATE: 1998-06-30
;   PRIOR APPLICATION NUMBER: US 60/123406
;   PRIOR FILING DATE: 1999-03-08
;   NUMBER OF SEQ ID NOS: 15
;   SOFTWARE: Patent.pm
; SEQ ID NO 1
;   LENGTH: 43069
;   TYPE: DNA
;   ORGANISM: homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: 1..7708
;   OTHER INFORMATION: potential 5'regulatory region
;   NAME/KEY: misc_feature
;   LOCATION: 36604..43069
;   OTHER INFORMATION: potential 3'regulatory region
;   NAME/KEY: exon
;   LOCATION: 7709..7852
;   OTHER INFORMATION: exon1
;   NAME/KEY: exon
;   LOCATION: 16236..16335
;   OTHER INFORMATION: exon2
;   NAME/KEY: exon
;   LOCATION: 24227..24297
;   OTHER INFORMATION: exon3
;   NAME/KEY: exon
;   LOCATION: 28133..28214
;   OTHER INFORMATION: exon4
;   NAME/KEY: exon
;   LOCATION: 36128..36605
;   OTHER INFORMATION: exon5
;   NAME/KEY: misc_feature
;   LOCATION: 7783..7785
;   OTHER INFORMATION: ATG
;   NAME/KEY: misc_feature
;   LOCATION: 36288..36290
;   OTHER INFORMATION: stop : TAA
;   NAME/KEY: polyA_signal
;   LOCATION: 36581..36586
;   OTHER INFORMATION: AATAAA
;   NAME/KEY: misc_feature
;   LOCATION: 7008..8116
;   OTHER INFORMATION: homology with sequence in ref genbank : M60470
;   NAME/KEY: misc_feature
;   LOCATION: 15995..16549
;   OTHER INFORMATION: homology with sequence in ref genbank : M63259
;   NAME/KEY: misc_feature
;   LOCATION: 24059..24597
;   OTHER INFORMATION: homology with sequence in ref genbank : M63260
;   NAME/KEY: misc_feature
;   LOCATION: 27873..28412
;   OTHER INFORMATION: homology with sequence in ref genbank : M63261
;   NAME/KEY: misc_feature
;   LOCATION: 35977..36926
;   OTHER INFORMATION: homology with sequence in ref genbank : M63262
;   NAME/KEY: misc_feature
;   LOCATION: 7613
;   OTHER INFORMATION: diverging nucleotide deletion of a A in ref : M60470
;   NAME/KEY: misc_feature
;   LOCATION: 16347
;   OTHER INFORMATION: diverging nucleotide G in ref : M63259
;   NAME/KEY: misc_feature
;   LOCATION: 16348

OTHER INFORMATION: diverging nucleotide A in ref : M63259
NAME/KEY: misc_feature
LOCATION: 24060
OTHER INFORMATION: diverging nucleotide deletion of a G in ref : M63260
NAME/KEY: misc_feature
LOCATION: 24067
OTHER INFORMATION: diverging nucleotide deletion of a G in ref : M63260
NAME/KEY: misc_feature
LOCATION: 27903
OTHER INFORMATION: diverging nucleotide deletion of a C in ref : M63261
NAME/KEY: misc_feature
LOCATION: 28327
OTHER INFORMATION: diverging nucleotide deletion of a G in ref : M63261
NAME/KEY: misc_feature
LOCATION: 3851..4189
OTHER INFORMATION: 10-517
NAME/KEY: misc_feature
LOCATION: 4120..4390
OTHER INFORMATION: 10-518
NAME/KEY: misc_feature
LOCATION: 4373..4792
OTHER INFORMATION: 10-253
NAME/KEY: misc_feature
LOCATION: 4814..5043
OTHER INFORMATION: 10-499
NAME/KEY: misc_feature
LOCATION: 4956..5422
OTHER INFORMATION: 10-500
NAME/KEY: misc_feature
LOCATION: 5524..5996
OTHER INFORMATION: 10-522
NAME/KEY: misc_feature
LOCATION: 6218..6672
OTHER INFORMATION: 10-503
NAME/KEY: misc_feature
LOCATION: 6522..6790
OTHER INFORMATION: 10-504
NAME/KEY: misc_feature
LOCATION: 7120..7574
OTHER INFORMATION: 10-204
NAME/KEY: misc_feature
LOCATION: 7513..7933
OTHER INFORMATION: 10-32
NAME/KEY: misc_feature
LOCATION: 16114..16533
OTHER INFORMATION: 10-33
NAME/KEY: misc_feature
LOCATION: 24072..24425
OTHER INFORMATION: 10-34
NAME/KEY: misc_feature
LOCATION: 27978..28401
OTHER INFORMATION: 10-35
NAME/KEY: misc_feature
LOCATION: 36020..36465
OTHER INFORMATION: 10-36
NAME/KEY: misc_feature
LOCATION: 16318..36669
OTHER INFORMATION: 10-498
NAME/KEY: misc_feature
LOCATION: 38441..38840
OTHER INFORMATION: 12-629
NAME/KEY: misc_feature
LOCATION: 42233..42749
OTHER INFORMATION: 12-628 complement
NAME/KEY: allele
LOCATION: 3950
OTHER INFORMATION: 10-517-100 : polymorphic base S
NAME/KEY: allele
LOCATION: 4243
OTHER INFORMATION: 10-518-125 : polymorphic base K
NAME/KEY: allele
LOCATION: 4312
OTHER INFORMATION: 10-518-194 : polymorphic base R

NAME/KEY: allele
LOCATION: 4490
OTHER INFORMATION: 10-253-118 : polymorphic base R
NAME/KEY: allele
LOCATION: 4670
OTHER INFORMATION: 10-253-298 : polymorphic base S
NAME/KEY: allele
LOCATION: 4687
OTHER INFORMATION: 10-253-315 : polymorphic base Y
NAME/KEY: allele
LOCATION: 4968
OTHER INFORMATION: 10-499-155 : polymorphic base R
NAME/KEY: allele
LOCATION: 5140
OTHER INFORMATION: 10-500-185 : polymorphic base Y
NAME/KEY: allele
LOCATION: 5213
OTHER INFORMATION: 10-500-258 : polymorphic base K
NAME/KEY: allele
LOCATION: 5364
OTHER INFORMATION: 10-500-410 : polymorphic base R
NAME/KEY: allele
LOCATION: 5594
OTHER INFORMATION: 10-522-71 : polymorphic base R
NAME/KEY: allele
LOCATION: 6370
OTHER INFORMATION: 10-503-159 : polymorphic base K
NAME/KEY: allele
LOCATION: 6693
OTHER INFORMATION: 10-504-172 : polymorphic base W
NAME/KEY: allele
LOCATION: 6763
OTHER INFORMATION: 10-504-243 : polymorphic base M
NAME/KEY: allele
LOCATION: 7445
OTHER INFORMATION: 10-204-326 : polymorphic base R
NAME/KEY: allele
LOCATION: 7870
OTHER INFORMATION: 10-32-357 : polymorphic base M
NAME/KEY: allele
LOCATION: 16288
OTHER INFORMATION: 10-33-175 : polymorphic base Y
NAME/KEY: allele
LOCATION: 16347
OTHER INFORMATION: 10-33-234 : polymorphic base M
NAME/KEY: allele
LOCATION: 16383
OTHER INFORMATION: 10-33-270 : polymorphic base R
NAME/KEY: allele

Query Match   75.2%; Score 15.8; DB 4; Length 43069;
Best Local Similarity 89.5%; Pred. No. 1.2e+02;
Matches   17; Conservative   0; Mismatches   2; Indels   0; Gaps   0;

Qy   2 GGATTCCTAGAGGAAGGA 20
             |||||||| |||||||||
Db   15865 GGATTCCTTGAGGGAATGA 15847

RESULT 9
US-08-724-394A-16/c
; Sequence 16, Application US/08724394A
; Patent No. 5872237
; GENERAL INFORMATION:
; APPLICANT: Feder, John N.
; APPLICANT: Kronmal, Gregory S.
; APPLICANT: Lauer, Peter M.
; APPLICANT: Ruddy, David A.
; APPLICANT: Thomas, Winston
; APPLICANT: Tsuchihashi, Zenta
; APPLICANT: Wolff, Roger K.
; TITLE OF INVENTION: Megabase Transcript Map: No. 5872237el
; TITLE OF INVENTION: Sequences and Antibodies Thereto

NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/724,394A
FILING DATE: 01-OCT-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Fitts, Renee A.
REGISTRATION NUMBER: 35,136
REFERENCE/DOCKET NUMBER: 017957-000100
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-576-0200
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 3502 base pairs
TYPE: nucleic acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1..3502
OTHER INFORMATION: /note= "cDNA 32"
US-08-724-394A-16

Query Match   73.3%; Score 15.4; DB 2; Length 3502;
Best Local Similarity 94.1%; Pred. No. 1.3e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 1; Gaps 0;

```
Qy     1 GGGATTCCTAGAGGGAA 17
           |||||||||||||||||
Db  2924 GGGATTCCTAGAGAGAA 2908
```

RESULT 10
US-09-621-976-12965
; Sequence 12965, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 12965
; LENGTH: 288
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 286
; OTHER INFORMATION: n=a, g, c or t
US-09-621-976-12965

Query Match   72.4%; Score 15.2; DB 4; Length 288;
Best Local Similarity 85.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy     2 GGATTCCTAGAGGGAAGAGAG 21
           ||||||||||| |||| ||||
Db   207 GGATTCCTGAGGTCAGGAG 226
```

RESULT 11
US-09-621-976-10679/c
; Sequence 10679, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 10679
; LENGTH: 351
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-10679

Query Match   72.4%; Score 15.2; DB 4; Length 351;
Best Local Similarity 85.0%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy     1 GGGATTCCTAGAGGGAAGGA 20
           ||| |||||—||||||
Db   318 GGGTTTCCTAGCGAGAAGGA 299
```

RESULT 12
US-09-621-976-2535/c
; Sequence 2535, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 2535
; LENGTH: 463
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 232..402
US-09-621-976-2535

Query Match   72.4%; Score 15.2; DB 4; Length 463;
Best Local Similarity 85.0%; Pred. No. 1.2e+02;
Matches 17; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
Qy     1 GGGATTCCTAGAGGGAAGGA 20
           ||||||||| — | ||||||
Db   436 GGGTTTCCTAGCGAGAAGGA 417
```

RESULT 13
US-09-621-976-18505/c
; Sequence 18505, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET 054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 18505
; LENGTH: 554
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-621-976-18505

Query Match   72.4%;   Score 15.2; DB 4;   Length 554;
Best Local Similarity 85.0%;   Pred. No. 1.2e+02;
Matches   17; Conservative   0; Mismatches 3;   Indels   0;   Gaps   0;

Qy   2 GGATTCCTAGAGGGAAGGAG 21
          |||||| || ||||||||
Db   543 GGATTCCAGGAGGAAGGAGGAG 524

RESULT 14
US-09-252-991A-13569/c
; Sequence 13569, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 13569
; LENGTH: 642
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-13569

Query Match   72.4%;   Score 15.2; DB 4;   Length 642;
Best Local Similarity 85.0%;   Pred. No. 1.2e+02;
Matches   17; Conservative   0; Mismatches 3;   Indels   0;   Gaps   0;

Qy   1 GGGATTCCTAGAGGGAAGGA 20
          |||||| || |||||||
Db   24 GGGATGCCTGGAGTGAAGGA 5

RESULT 15
US-09-252-991A-12999/c
; Sequence 12999, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12999
; LENGTH: 810
; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12999

Query Match   72.4%;   Score 15.2; DB 4;   Length 810;
Best Local Similarity 85.0%;   Pred. No. 1.3e+02;
Matches   17; Conservative   0; Mismatches 3;   Indels   0;   Gaps   0;

Qy   1 GGGATTCCTAGAGGGAAGGA 20
          |||||| || |||||||
Db   799 GGGTTTCCTTGAGGAAAGGA 780

Search completed: March 25, 2004, 15:34:29
Job time : 6.13061 secs

OM nucleic - nucleic search, using sw model

Copyright   GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: March 25, 2004, 08:39:03 ; Search time 61.3218 Seconds
                                          (without alignments)
                                          6361.316 Million cell updates/sec

Title: US-09-963-285-1_COPY_223_231
Perfect score: 9
Sequence: 1 acaaatgtt 9

Scoring table: IDENTITY_NUC
              Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 2167516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   GenEmbl:*
              1: gb_ba:*
              2: gb_htg:*
              3: gb_in:*
              4: gb_om:*
              5: gb_ov:*
              6: gb_pat:*
              7: gb_ph:*
              8: gb_pl:*
              9: gb_pr:*
             10: gb_ro:*
             11: gb_sts:*
             12: gb_sy:*
             13: gb_un:*
             14: gb_vi:*
             15: em_ba:*
             16: em_fun:*
             17: em_hum:*
             18: em_in:*
             19: em_mu:*
             20: em_om:*
             21: em_or:*
             22: em_ov:*
             23: em_pat:*
             24: em_ph:*
             25: em_pl:*
             26: em_ro:*
             27: em_sts:*
             28: em_un:*
             29: em_vi:*
             30: em_htg_hum:*
             31: em_htg_inv:*
             32: em_htg_other:*
             33: em_htg_mus:*
             34: em_htg_pln:*
             35: em_htg_rod:*
             36: em_htg_mam:*
             37: em_htg_vrt:*
             38: em_sy:*
             39: em_htgo_hum:*
             40: em_htgo_mus:*
             41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.		Score	% Query Match	Length	DB	ID	Description
1	c	9	100.0	9	17 6	AR039289	AR039289 Sequence
2	c	9	100.0	9	17 6	AR039291	AR039291 Sequence
3	c	9	100.0	9	17 6	AR040261	AR040261 Sequence
4	c	9	100.0	9	17 6	AR040263	AR040263 Sequence
5	c	9	100.0	9	17 6	BD256634	BD256634 Regulatio
6	c	9	100.0	9	17 6	BD256635	BD256635 Regulatio
7	c	9	100.0	9	17 6	BD257082	BD257082 Regulatio
8	c	9	100.0	9	17 6	BD257083	BD257083 Regulatio
9	c	9	100.0	9	17 6	AR190012	AR190012 Sequence
10	c	9	100.0	9	17 6	AR190013	AR190013 Sequence
11	c	9	100.0	9	17 6	AR324989	AR324989 Sequence
12	c	9	100.0	9	17 6	AR324990	AR324990 Sequence
13	c	9	100.0	9	17 6	AX674725	AX674725 Sequence
14	c	9	100.0	9	17 6	AX737260	AX737260 Sequence
15	c	9	100.0	9	17 6	AX738366	AX738366 Sequence
16	c	9	100.0	9	17 6	AX759329	AX759329 Sequence
17	c	9	100.0	9	17 6	AX761772	AX761772 Sequence
18	c	9	100.0	9	18 6	BD261376	BD261376 Cadherin-
19	c	9	100.0	9	18 6	AX353253	AX353253 Sequence
20	c	9	100.0	9	18 6	AX353256	AX353256 Sequence
21	c	9	100.0	9	18 6	AX353258	AX353258 Sequence
22	c	9	100.0	9	18 6	AX353267	AX353267 Sequence
23	c	9	100.0	9	18 6	AX353272	AX353272 Sequence
24	c	9	100.0	9	18 6	AX363098	AX363098 Sequence
25	c	9	100.0	9	18 6	AX363101	AX363101 Sequence
26	c	9	100.0	9	18 6	AX363103	AX363103 Sequence
27	c	9	100.0	9	18 6	AX363112	AX363112 Sequence
28	c	9	100.0	9	18 6	AX363117	AX363117 Sequence
29	c	9	100.0	9	19 6	AX130438	AX130438 Sequence
30	c	9	100.0	9	19 6	AX130439	AX130439 Sequence
31	c	9	100.0	9	19 6	AX130440	AX130440 Sequence
32	c	9	100.0	9	19 6	AX130441	AX130441 Sequence
33	c	9	100.0	9	20 6	AR224678	AR224678 Sequence
34	c	9	100.0	9	20 6	AR224679	AR224679 Sequence
35	c	9	100.0	9	20 6	AX392100	AX392100 Sequence
36	c	9	100.0	9	20 6	AX535426	AX535426 Sequence
37	c	9	100.0	9	20 6	AX823621	AX823621 Sequence
38	c	9	100.0	9	22 6	AR024478	AR024478 Sequence
39	c	9	100.0	9	22 6	AR089048	AR089048 Sequence
40	c	9	100.0	9	22 6	AR089481	AR089481 Sequence
41	c	9	100.0	9	22 6	AR140684	AR140684 Sequence
42	c	9	100.0	9	22 6	AR428284	AR428284 Sequence
43	c	9	100.0	9	22 6	AX352505	AX352505 Sequence
44	c	9	100.0	9	22 6	BD023016	BD023016 Species-s
45	c	9	100.0	9	23 6	AX077239	AX077239 Sequence

ALIGNMENTS

RESULT 1
AR039289/c
LOCUS   AR039289   Sequence 137 from patent US 5807743.   17 bp   DNA   linear   PAT 29-SEP-1999
DEFINITION Sequence 137 from patent US 5807743.
ACCESSION   AR039289
VERSION   AR039289.1   GI:5958652
KEYWORDS   .
SOURCE   Unknown.
  ORGANISM Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 17)
  AUTHORS   Stinchcomb,D.T. and McSwiggen,J.A.
  TITLE   Interleukin-2 receptor gamma-chain ribozymes
  JOURNAL   Patent: US 5807743-A 137 15-SEP-1998;
FEATURES   Location/Qualifiers

Pred. No. is the number of results predicted by chance to have a

```
AR040263/c
LOCUS       AR040263                    17 bp   DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1111 from patent US 5807743.
ACCESSION   AR040263
VERSION     AR040263.1  GI:5959626
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 17)
AUTHORS     Stinchcomb,D.T. and McSwiggen,J.A.
TITLE       Interleukin-2 receptor gamma-chain ribozymes
JOURNAL     Patent: US 5807743-A 1111 15-SEP-1998;
FEATURES    Location/Qualifiers
source      1..17
            /organism="unknown"
            /mol_type="unassigned DNA"
ORIGIN

Query Match           100.0%;  Score 9;  DB 6;  Length 17;
Best Local Similarity 100.0%;  Pred. No. 1.3e+05;
Matches      9;  Conservative   0;  Mismatches   0;  Indels    0;  Gaps    0;

Qy      1 ACAAATGTT 9
           |||||||||
Db     11 ACAAATGTT 3

RESULT 5
BD256634/c
LOCUS       BD256634                    17 bp   DNA     linear   PAT 17-JUL-2003
DEFINITION  Regulation of repressor genes using nucleic acid molecules.
ACCESSION   BD256634
VERSION     BD256634.1  GI:33066404
KEYWORDS    JP 2002541795-A/4427.
SOURCE      unidentified
ORGANISM    unidentified
            unclassified.
REFERENCE   1  (bases 1 to 17)
AUTHORS     Blatt,L., Zwick,M., Pavco,P. and McSwiggen,J.
TITLE       Regulation of repressor genes using nucleic acid molecules
JOURNAL     Patent: JP 2002541795-A 4427 10-DEC-2002;
            RIBOZYME PHARMACEUTICALS INC
COMMENT     OS   Eukaryote
            PN   JP 2002541795-A/4427
            PD   10-DEC-2002
            PF   11-APR-2000 JP 2000611654
            PR   12-APR-1999 US   60/129390
            PI   LAWRENCE BLATT,MICHAEL ZWICK,PAMELA PAVCO,JAMES MCSWIGGEN PC
            C12N5/09,A61K38/00,A61K48/00,A61P43/00,C12N5/10, PC
            C12P21/02,
            PC
            C12P21/02,C12P21/02//A61K31/711,(C12N5/10,C12R1:91),(C12P21/02, PC
            C12R1:91),
            PC   (C12P21/02,C12R1:91),(C12P21/02,C12R1:91),C12N15/00,C12N5/00,
            PC   A61K37/02,
            PC   (C12N5/00,C12R1:91)
            CC   Regulation of repressor genes using nucleic acid molecules FH
            Key           Location/Qualifiers
            FT   source        1..17
            FT                  /organism='Eukaryote'.
FEATURES    Location/Qualifiers
source      1..17
            /organism="unidentified"
            /mol_type="genomic DNA"
            /db_xref="taxon:32644"
ORIGIN

Query Match           100.0%;  Score 9;  DB 6;  Length 17;
Best Local Similarity 100.0%;  Pred. No. 1.3e+05;
Matches      9;  Conservative   0;  Mismatches   0;  Indels    0;  Gaps    0;

Qy      1 ACAAATGTT 9
```

```
source      1..17
            /organism="unknown"
            /mol_type="unassigned DNA"
ORIGIN

Query Match           100.0%;  Score 9;  DB 6;  Length 17;
Best Local Similarity 100.0%;  Pred. No. 1.3e+05;
Matches      9;  Conservative   0;  Mismatches   0;  Indels    0;  Gaps    0;

Qy      1 ACAAATGTT 9
           |||||||||
Db     12 ACAAATGTT 4

RESULT 2
AR039291/c
LOCUS       AR039291                    17 bp   DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 139 from patent US 5807743.
ACCESSION   AR039291
VERSION     AR039291.1  GI:5958654
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 17)
AUTHORS     Stinchcomb,D.T. and McSwiggen,J.A.
TITLE       Interleukin-2 receptor gamma-chain ribozymes
JOURNAL     Patent: US 5807743-A 139 15-SEP-1998;
FEATURES    Location/Qualifiers
source      1..17
            /organism="unknown"
            /mol_type="unassigned DNA"
ORIGIN

Query Match           100.0%;  Score 9;  DB 6;  Length 17;
Best Local Similarity 100.0%;  Pred. No. 1.3e+05;
Matches      9;  Conservative   0;  Mismatches   0;  Indels    0;  Gaps    0;

Qy      1 ACAAATGTT 9
           |||||||||
Db     11 ACAAATGTT 3

RESULT 3
AR040261/c
LOCUS       AR040261                    17 bp   DNA     linear   PAT 29-SEP-1999
DEFINITION  Sequence 1109 from patent US 5807743.
ACCESSION   AR040261
VERSION     AR040261.1  GI:5959624
KEYWORDS    .
SOURCE      Unknown.
ORGANISM    Unknown.
            Unclassified.
REFERENCE   1  (bases 1 to 17)
AUTHORS     Stinchcomb,D.T. and McSwiggen,J.A.
TITLE       Interleukin-2 receptor gamma-chain ribozymes
JOURNAL     Patent: US 5807743-A 1109 15-SEP-1998;
FEATURES    Location/Qualifiers
source      1..17
            /organism="unknown"
            /mol_type="unassigned DNA"
ORIGIN

Query Match           100.0%;  Score 9;  DB 6;  Length 17;
Best Local Similarity 100.0%;  Pred. No. 1.3e+05;
Matches      9;  Conservative   0;  Mismatches   0;  Indels    0;  Gaps    0;

Qy      1 ACAAATGTT 9
           |||||||||
Db     12 ACAAATGTT 4

RESULT 4
```

Db   12 ACAAATGTT 4
               |||||||||

RESULT 6
BD256635/c
LOCUS   BD256635   17 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION Regulation of repressor genes using nucleic acid molecules.
ACCESSION   BD256635
VERSION   BD256635.1 GI:33066405
KEYWORDS   JP 2002541795-A/4428.
SOURCE   unidentified
ORGANISM   unidentified
            unclassified.
REFERENCE   1 (bases 1 to 17)
AUTHORS   Blatt,L., Zwick,M., Pavco,P. and Mcswiggen,J.
TITLE   Regulation of repressor genes using nucleic acid molecules
JOURNAL   Patent: JP 2002541795-A 4428 10-DEC-2002;
            RIBOZYME PHARMACEUTICALS INC
COMMENT   OS Eukaryote
            PN JP 2002541795-A/4428
            PD 10-DEC-2002
            PF 11-APR-2000 JP 2000611654
            PR 12-APR-1999 US   60/129390
            PI LAWRENCE BLATT,MICHAEL ZWICK,PAMELA PAVCO,JAMES MCSWIGGEN PC
            C12N15/09,A61K38/00,A61K48/00,A61P43/00,C12N5/10, PC
            C12P21/02,
            PC
            C12P21/02,C12P21/02//A61K31/711,(C12P21/02,C12R1:91),(C12P21/02, PC
            C12R1:91),
            PC A61K37/02,(C12P21/02,C12R1:91),(C12P21/02,C12R1:91),C12N15/00,C12N5/00,
            PC (C12N5/00,C12R1:91)
            CC Regulation of repressor genes using nucleic acid molecules FH
            Key   Location/Qualifiers
            FT source   1..17
            FT   /organism='Eukaryote'.
FEATURES   Location/Qualifiers
     source   1..17
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

Query Match   100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches   9; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 ACAAATGTT 9
            |||||||||
Db   11 ACAAATGTT 3

RESULT 7
BD257082/c
LOCUS   BD257082   17 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION Regulation of repressor genes using nucleic acid molecules.
ACCESSION   BD257082
VERSION   BD257082.1 GI:33066852
KEYWORDS   JP 2002541795-A/4875.
SOURCE   unidentified
ORGANISM   unidentified
            unclassified.
REFERENCE   1 (bases 1 to 17)
AUTHORS   Blatt,L., Zwick,M., Pavco,P. and Mcswiggen,J.
TITLE   Regulation of repressor genes using nucleic acid molecules
JOURNAL   Patent: JP 2002541795-A 4875 10-DEC-2002;
            RIBOZYME PHARMACEUTICALS INC
COMMENT   OS Eukaryote
            PN JP 2002541795-A/4875
            PD 10-DEC-2002
            PF 11-APR-2000 JP 2000611654
            PR 12-APR-1999 US   60/129390

PI LAWRENCE BLATT,MICHAEL ZWICK,PAMELA PAVCO,JAMES MCSWIGGEN PC
C12N15/09,A61K38/00,A61K48/00,A61P43/00,C12N5/10,C12N5/00, PC
C12P21/02,
PC
C12P21/02,C12P21/02//A61K31/711,(C12N5/10,C12R1:91),(C12P21/02, PC
C12R1:91),
PC (C12P21/02,C12R1:91),(C12P21/02,C12R1:91),C12N15/00,C12N5/00,
PC A61K37/02,
PC (C12N5/00,C12R1:91)
CC Regulation of repressor genes using nucleic acid molecules FH
Key   Location/Qualifiers
FT source   1..17
FT   /organism='Eukaryote'.
FEATURES   Location/Qualifiers
     source   1..17
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

Query Match   100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches   9; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 ACAAATGTT 9
            |||||||||
Db   12 ACAAATGTT 4

RESULT 8
BD257083/c
LOCUS   BD257083   17 bp   DNA   linear   PAT 17-JUL-2003
DEFINITION Regulation of repressor genes using nucleic acid molecules.
ACCESSION   BD257083
VERSION   BD257083.1 GI:33066853
KEYWORDS   JP 2002541795-A/4876.
SOURCE   unidentified
ORGANISM   unidentified
            unclassified.
REFERENCE   1 (bases 1 to 17)
AUTHORS   Blatt,L., Zwick,M., Pavco,P. and Mcswiggen,J.
TITLE   Regulation of repressor genes using nucleic acid molecules
JOURNAL   Patent: JP 2002541795-A 4876 10-DEC-2002;
            RIBOZYME PHARMACEUTICALS INC
COMMENT   OS Eukaryote
            PN JP 2002541795-A/4876
            PD 10-DEC-2002
            PF 11-APR-2000 JP 2000611654
            PR 12-APR-1999 US   60/129390
            PI LAWRENCE BLATT,MICHAEL ZWICK,PAMELA PAVCO,JAMES MCSWIGGEN PC
            C12N15/09,A61K38/00,A61K48/00,A61P43/00,C12N5/10, PC
            C12P21/02,
            PC
            C12P21/02,C12P21/02//A61K31/711,(C12N5/10,C12R1:91),(C12P21/02, PC
            C12R1:91),
            PC A61K37/02,(C12P21/02,C12R1:91),(C12P21/02,C12R1:91),C12N15/00,C12N5/00,
            PC (C12N5/00,C12R1:91)
            CC Regulation of repressor genes using nucleic acid molecules FH
            Key   Location/Qualifiers
            FT source   1..17
            FT   /organism='Eukaryote'.
FEATURES   Location/Qualifiers
     source   1..17
                     /organism="unidentified"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:32644"
ORIGIN

Query Match   100.0%; Score 9; DB 6; Length 17;
Best Local Similarity 100.0%; Pred. No. 1.3e+05;
Matches   9; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    1 ACAAATGTT 9
        |||||||||
Db   11 ACAAATGTT 3
```

RESULT 9
AR190012/c
LOCUS   AR190012   17 bp   DNA   linear   PAT 20-APR-2002
DEFINITION Sequence 5500 from patent US 6346398.
ACCESSION   AR190012
VERSION   AR190012.1 GI:20235977
KEYWORDS
SOURCE   Unknown.
ORGANISM   Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 17)
AUTHORS   Pavco,P., McSwiggen,J., Stinchcomb,D. and Escobedo,J.
TITLE   Method and reagent for the treatment of diseases or conditions
            related to levels of vascular endothelial growth factor receptor
JOURNAL   Patent: US 6346398-A 5500 12-FEB-2002;
FEATURES   Location/Qualifiers
            source   1..17
                      /organism="unknown"
                      /mol_type="unassigned DNA"
ORIGIN

Query Match   100.0%; Score 9; DB 6; Length 17;
Best Local Similarity   100.0%; Pred. No. 1.3e+05;
Matches   9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 ACAAATGTT 9
        |||||||||
Db   12 ACAAATGTT 4
```

RESULT 10
AR190013/c
LOCUS   AR190013   17 bp   DNA   linear   PAT 20-APR-2002
DEFINITION Sequence 5501 from patent US 6346398.
ACCESSION   AR190013
VERSION   AR190013.1 GI:20235978
KEYWORDS
SOURCE   Unknown.
ORGANISM   Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 17)
AUTHORS   Pavco,P., McSwiggen,J., Stinchcomb,D. and Escobedo,J.
TITLE   Method and reagent for the treatment of diseases or conditions
            related to levels of vascular endothelial growth factor receptor
JOURNAL   Patent: US 6346398-A 5501 12-FEB-2002;
FEATURES   Location/Qualifiers
            source   1..17
                      /organism="unknown"
                      /mol_type="unassigned DNA"
ORIGIN

Query Match   100.0%; Score 9; DB 6; Length 17;
Best Local Similarity   100.0%; Pred. No. 1.3e+05;
Matches   9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 ACAAATGTT 9
        |||||||||
Db   12 ACAAATGTT 4
```

RESULT 11
AR324989/c
LOCUS   AR324989   17 bp   RNA   linear   PAT 17-AUG-2003
DEFINITION Sequence 2391 from patent US 6566127.
ACCESSION   AR324989
VERSION   AR324989.1 GI:33710797
KEYWORDS
SOURCE   Unknown.
ORGANISM   Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 17)
AUTHORS   Pavco,P., McSwiggen,J.A., Stinchcomb,D.T. and Escobedo,J.
TITLE   Method and reagent for the treatment of diseases or conditions
            related to levels of vascular endothelial growth factor receptor
JOURNAL   Patent: US 6566127-A 2391 20-MAY-2003;
FEATURES   Location/Qualifiers
            source   1..17
                      /organism="unknown"
                      /mol_type="unassigned RNA"
ORIGIN

Query Match   100.0%; Score 9; DB 6; Length 17;
Best Local Similarity   100.0%; Pred. No. 1.3e+05;
Matches   9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 ACAAATGTT 9
        |||||||||
Db   12 ACAAATGTT 4
```

RESULT 12
AR324990/c
LOCUS   AR324990   17 bp   RNA   linear   PAT 17-AUG-2003
DEFINITION Sequence 2392 from patent US 6566127.
ACCESSION   AR324990
VERSION   AR324990.1 GI:33710798
KEYWORDS
SOURCE   Unknown.
ORGANISM   Unknown.
            Unclassified.
REFERENCE   1 (bases 1 to 17)
AUTHORS   Pavco,P., McSwiggen,J.A., Stinchcomb,D.T. and Escobedo,J.
TITLE   Method and reagent for the treatment of diseases or conditions
            related to levels of vascular endothelial growth factor receptor
JOURNAL   Patent: US 6566127-A 2392 20-MAY-2003;
FEATURES   Location/Qualifiers
            source   1..17
                      /organism="unknown"
                      /mol_type="unassigned RNA"
ORIGIN

Query Match   100.0%; Score 9; DB 6; Length 17;
Best Local Similarity   100.0%; Pred. No. 1.3e+05;
Matches   9; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 ACAAATGTT 9
        |||||||||
Db   11 ACAAATGTT 3
```

RESULT 13
AX674725
LOCUS   AX674725   17 bp   DNA   linear   PAT 27-MAR-2003
DEFINITION Sequence 3170 from patent WO03004526.
ACCESSION   AX674725
VERSION   AX674725.1 GI:29333073
KEYWORDS
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Telerman,A., Amson,R. and Tuijnder,M.
TITLE   Sequences involved in phenomena of tumour suppression, tumour
            reversion, apoptosis and/or resistance to viruses and their use as
            medicines
JOURNAL   Patent: WO 03004526-A 3170 16-JAN-2003;
            Molecular Engines Laboratories (FR)
FEATURES   Location/Qualifiers
            source   1..17
                      /organism="Homo sapiens"

ORIGIN
```
                /mol_type="unassigned DNA"
                /db_xref="taxon:9606"

Query Match           100.0%;  Score 9;  DB 6;  Length 17;
Best Local Similarity 100.0%;  Pred. No. 1.3e+05;
Matches        9;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy       1 ACAAATGTT 9
           |||||||||
Db       9 ACAAATGTT 17
```

RESULT 14
AX737260/c
LOCUS   AX737260   17 bp   DNA   linear   PAT 08-MAY-2003
DEFINITION Sequence 2850 from Patent WO03025177.
ACCESSION   AX737260
VERSION   AX737260.1 GI:30516548
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Telerman,A., Amson,R. and Tuijnder,M.
TITLE   Sequences involved in phenomena of tumour suppression, tumour
            reversion, apoptosis and/or resistance to viruses and the use
            thereof as medicaments
JOURNAL   Patent: WO 03025177-A 2850 27-MAR-2003;
            Molecular Engines Laboratories (FR)
FEATURES   Location/Qualifiers
     source   1..17
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

```
Query Match           100.0%;  Score 9;  DB 6;  Length 17;
Best Local Similarity 100.0%;  Pred. No. 1.3e+05;
Matches        9;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy       1 ACAAATGTT 9
           |||||||||
Db      17 ACAAATGTT 9
```

RESULT 15
AX738366
LOCUS   AX738366   17 bp   DNA   linear   PAT 08-MAY-2003
DEFINITION Sequence 3956 from Patent WO03025177.
ACCESSION   AX738366
VERSION   AX738366.1 GI:30517654
KEYWORDS   .
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Telerman,A., Amson,R. and Tuijnder,M.
TITLE   Sequences involved in phenomena of tumour suppression, tumour
            reversion, apoptosis and/or resistance to viruses and the use
            thereof as medicaments
JOURNAL   Patent: WO 03025177-A 3956 27-MAR-2003;
            Molecular Engines Laboratories (FR)
FEATURES   Location/Qualifiers
     source   1..17
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
ORIGIN

Query Match   100.0%; Score 9; DB 6; Length 17;
```

Best Local Similarity 100.0%;  Pred. No. 1.3e+05;
Matches        9;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

```
Qy       1 ACAAATGTT 9
           |||||||||
Db       9 ACAAATGTT 17
```

Search completed: March 25, 2004, 12:49:45
Job time : 64.3218 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:　March 25, 2004, 08:39:03 ; Search time 1369.52 Seconds
　　　　　　　　　　　　　　　　　　　(without alignments)
　　　　　　　　　　　　　　　　　　　6361.316 Million cell updates/sec

Title:　　　　　US-09-963-285-1_COPY_223_423
Perfect score:　201
Sequence:　　　1 acaaatgttcctgaagc.........ggattcctagaggaggag 201

Scoring table:　IDENTITY_NUC
　　　　　　　　Gapop 10.0 , Gapext 1.0

Searched:　3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters:　6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
　　　　　　　　 Maximum Match 100%
　　　　　　　　 Listing first 45 summaries

Database :　GenEmbl:*
　1: gb_ba:*
　2: gb_htg:*
　3: gb_in:*
　4: gb_om:*
　5: gb_ov:*
　6: gb_pat:*
　7: gb_ph:*
　8: gb_pl:*
　9: gb_pr:*
　10: gb_ro:*
　11: gb_sts:*
　12: gb_sy:*
　13: gb_un:*
　14: gb_vi:*
　15: em_ba:*
　16: em_fun:*
　17: em_hum:*
　18: em_in:*
　19: em_mu:*
　20: em_om:*
　21: em_or:*
　22: em_ov:*
　23: em_pat:*
　24: em_ph:*
　25: em_pl:*
　26: em_ro:*
　27: em_sts:*
　28: em_un:*
　29: em_vi:*
　30: em_htg_hum:*
　31: em_htg_inv:*
　32: em_htg_other:*
　33: em_htg_mus:*
　34: em_htg_pln:*
　35: em_htg_rod:*
　36: em_htg_mam:*
　37: em_htg_vrt:*
　38: em_sy:*
　39: em_htgo_hum:*
　40: em_htgo_mus:*
　41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	201	100.0	168656	9	AC009108	AC009108 Homo sapi
2	66	32.8	6021	10	MMMFHEAD1	Y08222 M.musculus
3	66	32.8	178416	10	AC127554	AC127554 Mus muscu
c 4	38	18.9	186021	9	AC136621	AC136621 Homo sapi
5	38	18.9	188923	2	AC026802	AC026802 Homo sapi
c 6	36.8	18.3	129727	9	AC084381	AC084381 Homo sapi
7	36.8	18.3	241394	2	AC131549	AC131549 Rattus no
8	36.4	18.3	261236	2	AC119381	AC119381 Rattus no
9	36.4	18.1	128680	2	AC023374	AC023374 Homo sapi
c 10	36.4	18.1	146376	9	AC009247	AC009247 Homo sapi
c 11	36.4	18.1	166236	2	AC022832	AC022832 Homo sapi
c 12	35.2	17.5	163579	2	AC104476	AC104476 Pan trogl
c 13	35.2	17.5	167671	5	BX000462	BX000462 Zebrafish
14	35.2	17.5	180971	5	AC079310	AC079310 Homo sapi
15	35.2	17.5	189932	2	BX649492	BX649492 Danio rer
16	35.2	17.5	195516	5	AC053540	AC053540 Homo sapi
17	35.2	17.5	273859	5	BX649333	BX649333 Zebrafish
18	35	17.4	137271	9	AC004552	AC004552 Homo sapi
19	34.8	17.3	174437	9	AC018470	AC018470 Homo sapi
20	34.6	17.2	96887	8	AC066689	AC066689 Arabidops
21	34.4	17.1	158151	2	AC103904	AC103904 Canis fam
c 22	34.4	17.1	161338	2	AC105372	AC105372 Canis fam
23	34.4	17.1	162407	2	AC073969	AC073969 Homo sapi
24	34.4	17.1	187543	9	AC068130	AC068130 Homo sapi
25	34	16.9	176345	4	AC104675	AC104675 Sus scrof
26	34	16.9	225962	2	AC094695	AC094695 Rattus no
c 27	34	16.9	231904	2	AC114208	AC114208 Rattus no
28	34	16.9	232945	2	AC118407	AC118407 Rattus no
29	34	16.9	237544	2	AC109858	AC109858 Rattus no
30	33.8	16.8	5062	5	BMU251223	AJ251223 Bungarus
31	33.8	16.8	151482	9	AC068765	AC068765 Homo sapi
32	33.8	16.8	163535	2	AC026876	AC026876 Homo sapi
33	33.8	16.8	189476	2	AC009893	AC009893 Homo sapi
34	33.6	16.8	339168	2	AC078987	AC078987 Homo sapi
35	33.6	16.7	193795	2	AC139167	AC139167 Rattus no
36	33.6	16.7	261260	2	AC110942	AC110942 Rattus no
37	33.6	16.7	314578	2	AC115349	AC115349 Rattus no
c 38	33.4	16.6	153133	9	AC091980	AC091980 Homo sapi
39	33.4	16.6	175154	9	AC093246	AC093246 Homo sapi
c 40	33.2	16.5	100310	2	F17A13	AL096692 Arabidops
41	33.2	16.5	100984	10	AL929138	AL929138 Mouse DNA
42	33.2	16.5	162564	10	AC127575	AC127575 Mus muscu
43	33.2	16.5	181829	9	AC023425	AC023425 Homo sapi
c 44	33.2	16.5	184091	2	BX294393	BX294393 Mus muscu
c 45	33.2	16.5	197248	2	AC134907	AC134907 Mus muscu

ALIGNMENTS

RESULT 1
AC009108/c
LOCUS　　　　AC009108　　　168656 bp　　　DNA　　　linear　　PRI 29-OCT-2002
DEFINITION　Homo sapiens chromosome 16 clone RP11-46309, complete sequence.
ACCESSION　 AC009108
VERSION　　 AC009108.10　GI:24418066
KEYWORDS　　HTG.
SOURCE　　　Homo sapiens (human)
　ORGANISM　Homo sapiens
　　　　　　Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
　　　　　　Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE　 1　(bases 1 to 168656)
　AUTHORS　 DOE Joint Genome Institute, Stanford Human Genome Center and Los
　　　　　　Alamos National Laboratory.
　TITLE　　 Direct Submission

Pred. No. is the number of results predicted by chance to have a

JOURNAL   Medicine, Department of Biochemistry, 1-1-1 Hondo, Akita 010, JAPAN
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 168656)
AUTHORS   DOE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint
             Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3 (bases 1 to 168656)
AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
             Alamos National Laboratory.
TITLE   Direct Submission
JOURNAL   Submitted (22-OCT-2002) DOE Joint Genome Institute, 2800 Mitchell
             Drive, Walnut Creek, CA 94598, USA
COMMENT   On Oct 29, 2002 this sequence version replaced gi:1378786306.
             Draft Sequence Produced by DOE Joint Genome Institute
             www.jgi.doe.gov
             Finishing Completed at Stanford Human Genome Center and Los Alamos
             National Laboratory
             www-shgc.stanford.edu
             Quality: Phrap Quality >=40 99.9% of Sequence;
             Estimated Total Number of Errors is 0.2.
FEATURES   Location/Qualifiers
     source   1..168656
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="RP11-46309"

ORIGIN

Query Match   100.0%;   Score 201;   DB 9;   Length 168656;
Best Local Similarity 100.0%;   Pred. No. 2.3e-54;
Matches 201; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 ACAAATGTTCTCCCTGAAGCCCTCTTCCCTGCCAACCAGACCAGCAACTTCCAAAATTC 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 78510 ACAAATGTTCTCCCTGAAGCCCTCTTCCCTGCCAACCAGACCAGCAACTTCCAAAATTC 78451

Qy   61 TGCCCGTGTTTAGCCTTGTTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAGG 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 78450 TGCCCGTGTTTAGCCTTGTTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAGG 78391

Qy   121 GATCTGATTATTGAGGTGTGGAAGGAATAAATCAGTCCACAAATAAACAAACTGTCC 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 78390 GATCTGATTATTGAGGTGTGGAAGGAATAAATCAGTCCACAAATAAACAAACTGTCC 78331

Qy   181 GGGATTCCTAGAGGGAAGGAG 201
          |||||||||||||||||||||
Db 78330 GGGATTCCTAGAGGGAAGGAG 78310

RESULT 2
MMMFHEAD1
LOCUS   MMMFHEAD1   6021 bp   DNA   linear   ROD 14-MAY-1997
DEFINITION M.musculus MFH-1 gene.
ACCESSION   Y08222
VERSION   Y08222.1 GI:1869968
KEYWORDS   mesenchyme fork head-1 protein; MFH-1 gene.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1
AUTHORS   Miura,N., Iida,K., Kakinuma,H., Yang,X.L. and Sugiyama,T.
TITLE   Isolation of the mouse (MFH-1) and human (FKHL 14) mesenchyme fork
            head-1 genes reveals conservation of their gene and protein
            structures
JOURNAL   Genomics 41 (3), 489-492 (1997)
MEDLINE   97312712
PUBMED   9169153
REFERENCE   2 (bases 1 to 6021)
AUTHORS   Miura,N.
TITLE   Direct Submission
JOURNAL   Submitted (18-SEP-1996) N. Miura, Akita University School of

FEATURES   Location/Qualifiers
     source   1..6021
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /strain="129"
                     /db_xref="taxon:10090"
     mRNA   1486..4345
                     /gene="MFH-1"
     gene   2070..3554
                     /gene="MFH-1"
     CDS   2070..3554
                     /gene="MFH-1"
                     /codon_start=1
                     /product="mesenchyme fork head-1 protein"
                     /protein_id="CAA69399.1"
                     /db_xref="GI:1869969"
                     /db_xref="GOA:Q61850"
                     /db_xref="SWISS-PROT:Q61850"
                     /translation="MQARYSVSDPNALGVVPYLSEQNYYRAAGSYGMASPMGVYSGH
                     PEQYGAGMGRSYAPYHQPAAPKDLVKPPYSYIALITMAIQNAPEKKITLNGIYQFIM
                     DRFPFYRENKQGWQNSIRHNLSLNECFVKVPRDDKKPGKGSYWTLDPDSYNMFENGSF
                     LRRRRKKKDVPKDKEERAHLKEPSTTAKGAPTGTPVADGPKEAEKKVVKSEAAS
                     PALPVITKVETLSPEGALQASPRSASSTPAGSPDGSLPEHAAAPNGLPGFSVETIMT
                     LRTSPPGGDLSPAAARAGLVVPPLALPYAAAPPAAYTQPCAQGLEAAGSAGYQCSMRA
                     MSLYTGAERPAHVCVPPALDEALSDHPSGPGSPGLAINLAAGQEGALGASGHHHQHHG
                     HLHPQAPPAPQPPPAPQPPATQATSWYLNHGGDLSHLPGHTFATQQTFPNVREMPNS
                     HRLGLDNSSLGESQVSNASCQLPYRATPSLVRHAAPYSYDCTKY"
     polyA_signal   4320..4325

ORIGIN

Query Match   32.8%;   Score 66;   DB 10;   Length 6021;
Best Local Similarity 73.1%;   Pred. No. 2.8e-10;
Matches 155; Conservative   0; Mismatches 45; Indels 12; Gaps   5;

Qy   1 ACAAATGTTCTCCCTGAAGCCCTCTTCCCTGCCAACCAGACCAGCAACTTCCAAAATTC 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   195 ACAAATGTTTTATCTGTCGCCTTCTTCCCTACCCACCCGGACCAACAACTTCCAGAAGGT 254

Qy   61 TGCCCGTG---TTTAGCCTTGTTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAA 117
          |||||||| |||| ||||| | ||||| ||||||||||||||||||||||||||||
Db   255 TCTGCGAGGCATAGAGCATTCCGTAGGACATCTCGGTGCTTCTGAGGAAGCGGGACCGA 314

Qy   118 G--GGGATCTGATTA----TTGAGGTGTGGAAGGAATAAATAAACCAGTCCACAAATAAAC 171
          | ||||||||||| |||||||||||||||||||||||| |||||||||||||||
Db   315 GCAGGGATCCGATGACGACTGAGATGTTGAAGGAATAAAT-ACCAGTCCACAAATAAAC 373

Qy   172 AAACTGT--CCGGGATTCCTAGAGGGAAGGAG 201
          ||||||| |||||||||||||||||||||||
Db   374 AAACTGTCCCCGGGATTCCTAGAGGGAAGGAG 405

RESULT 3
AC127554
LOCUS   AC127554   178416 bp   DNA   linear   ROD 27-NOV-2003
DEFINITION Mus musculus BAC clone RP24-323K23 from chromosome 8, complete
            sequence.
ACCESSION   AC127554
VERSION   AC127554.4 GI:33457241
KEYWORDS   HTG.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 178416)
AUTHORS   Cordes,M. and Haglund,K.
TITLE   The sequence of Mus musculus BAC clone RP24-323K23
JOURNAL   Unpublished (2001)
REFERENCE   2 (bases 1 to 178416)
AUTHORS   Wilson.R.
TITLE   Sequencing of Mus musculus
JOURNAL   Unpublished (2001)
REFERENCE   3 (bases 1 to 178416)
AUTHORS   McPherson,J.D. and Waterston,R.H.

```
TITLE      Direct Submission
JOURNAL    Submitted (17-JUL-2002) Genome Sequencing Center, 4444 Forest Park
           Parkway, St. Louis, MO 63108, USA
REFERENCE  4 (bases 1 to 178416)
AUTHORS    Wilson.R.K.
TITLE      Direct Submission
JOURNAL    Submitted (30-JUL-2003) Genome Sequencing Center, 4444 Forest Park
           Parkway, St. Louis, MO 63108, USA
REFERENCE  5 (bases 1 to 178416)
AUTHORS    Wilson.R.K.
TITLE      Direct Submission
JOURNAL    Submitted (06-AUG-2003) Genome Sequencing Center, 4444 Forest Park
           Parkway, St. Louis, MO 63108, USA
REFERENCE  6 (bases 1 to 178416)
AUTHORS    Wilson.R.
TITLE      Direct Submission
JOURNAL    Submitted (27-NOV-2003) Department of Genetics, Washington
           University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT    On Aug 6, 2003 this sequence version replaced gi:33342444.
           ------------- Genome Center
           Center: Washington University Genome Sequencing Center
           Center code: WUGSC
           Web site: http://genome.wustl.edu
           Contact: submissions@watson.wustl.edu
           ------------- Summary Statistics
           Center project name: M_BB0323K23
           -------------

NOTICE:    This sequence may not represent the entire insert of this
           clone.  It may be shorter because we only sequence overlapping
           clone sections once, or longer because we provide a small overlap
           between neighboring data submissions.

           This sequence was finished as follows unless otherwise noted:
           all regions were double stranded, sequenced with an alternate
           chemistry, or covered by high quality data (i.e., phred quality >=
           30); an attempt was made to resolve all sequencing problems, such
           as compressions and repeats; all regions were covered by sequence
           from more than one subclone; and the assembly was confirmed by
           restriction digest.

MAPPING INFORMATION:
           Mapping information for this clone was provided by Dr. Wes Warren,
           Department of Genetics, Washington University, St. Louis MO. For
           additional information about the map position of this sequence, see
           http://genome.wustl.edu

SOURCE INFORMATION:
           The RPCI-24 BAC Library has been constructed by Pieter de Jong and
           coworkers (http://www.chori.org) from male C57Bl/6J mouse spleen
           and/or brain genomic DNA. The clone and detailed information can be
           obtained from Pieter de Jong and coworkers at http://www.chori.org

NEIGHBORING SEQUENCE INFORMATION:
           This sequence is the entire insert of the clone. This clone is
           overlapped by AC124170.
FEATURES             Location/Qualifiers
     source          1..178416
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /chromosome="8"
                     /map="8"
                     /clone="RP24-323K23"
                     /clone_lib="RPCI-24"
     repeat_region   1791..1839
                     /rpt_family="Alu"
     repeat_region   3054..3288
                     /rpt_family="B4"
     repeat_region   3834..3845
                     /rpt_family="B4"
     repeat_region   3846..3972
                     /rpt_family="Alu"
     repeat_region   3973..4026
                     /rpt_family="B4"
     misc_feature    6269..8820
                     /note="CpG_island (%GC=69.2, o/e=0.78, #CpGs=269)"
     repeat_region   16848..16980
                     /rpt_family="MIR"
     repeat_region   18383..18494
                     /rpt_family="Alu"
     repeat_region   18974..19317
                     /rpt_family="MaLR"
     repeat_region   19821..20013
                     /rpt_family="B2"
     repeat_region   20188..20538
                     /rpt_family="MaLR"
     repeat_region   20539..20683
                     /rpt_family="Alu"
     repeat_region   20684..20799
                     /rpt_family="MaLR"
     repeat_region   22901..22957
                     /rpt_family="ID"
     repeat_region   22968..23090
                     /rpt_family="Alu"
     repeat_region   24384..24441
                     /rpt_family="B4"
     repeat_region   25838..26022
                     /rpt_family="MER1_type"
     repeat_region   27267..27380
                     /rpt_family="Alu"
     repeat_region   28700..28893
                     /rpt_family="B2"
     repeat_region   29950..30049
                     /rpt_family="B4"
     repeat_region   30809..30854
                     /rpt_family="ID"
     repeat_region   30859..30963
                     /rpt_family="Alu"
     repeat_region   30987..31062
                     /rpt_family="ID"
     repeat_region   31157..31294
                     /rpt_family="Alu"
     repeat_region   32192..32321
                     /rpt_family="B4"
     repeat_region   32923..33036
                     /rpt_family="Alu"
     repeat_region   33141..33245
                     /rpt_family="Alu"
     repeat_region   33394..33521
                     /rpt_family="B4"
     repeat_region   33806..33891
                     /rpt_family="B4"
     repeat_region   34643..34976
                     /rpt_family="MaLR"
     repeat_region   35547..35686
                     /rpt_family="B2"
     repeat_region   35701..35839
                     /rpt_family="Alu"
     repeat_region   35873..35982
                     /rpt_family="Alu"
     misc_feature    39796..41776
                     /note="CpG_island (%GC=69.3, o/e=0.84, #CpGs=225)"
     repeat_region   47196..47354
                     /rpt_family="B2"
     repeat_region   47365..47498
                     /rpt_family="Alu"
     repeat_region   47507..47550
                     /rpt_family="B4"
     repeat_region   47706..47840
                     /rpt_family="Alu"
     repeat_region   47716..47892
                     /rpt_family="B4"
     repeat_region   47952..48067
                     /rpt_family="Alu"
     misc_feature    51423..52857
```

repeat_region   /note="CpG island (%GC=66.5, o/e=0.72, #CpGs=121)"
53367..53484
repeat_region   /rpt_family="Alu"
54128..54233
repeat_region   /rpt_family="Alu"
55219..55240
repeat_region   /rpt_family="RVK"
55241..55364
repeat_region   /rpt_family="Alu"
55365..55460
repeat_region   /rpt_family="RVK"
56094..56379
repeat_region   /rpt_family="B4"
59534..59702
repeat_region   /rpt_family="B2"
60089..60165
repeat_region   /rpt_family="ID"
60513..60709
repeat_region   /rpt_family="B2"
60886..61042
repeat_region   /rpt_family="B4"
61852..61997
repeat_region   /rpt_family="MER1_type"
62247..62549
repeat_region   /rpt_family="MaLR"
62716..62764
repeat_region   /rpt_family="ID"
65656..65793
repeat_region   /rpt_family="Alu"
65929..66103
repeat_region   /rpt_family="MIR"
67019..67319
repeat_region   /rpt_family="MER121"
67684..67898

Query Match   32.8%;   Score 66;   DB 10;   Length 178416;
Best Local Similarity   73.1%;   Pred. No. 3.4e-10;
Matches 155;   Conservative   0;   Mismatches 45;   Indels 12;   Gaps 5;

Qy   1 ACAAATGTTCTCCCTGAAGCCCTCTTCCCTGCCCAACCAGACCACTTCCAAAATTC 60
            ||||||||||| |||| || ||||| ||||||||||||||| ||||||||||||
Db   38322 ACAAATGTTTATCTGTCGCCTTCTTCCCTACCCAACCAACTTCCAGAAGGT 38381

Qy   61 TGCCCGTG---TTTAGCCTTGTTAAAGGGGTGTCTCCACTCCTTCAGGGAAAGTGGGAAAA 117
            ||||||| || ||| ||| ||||| |||||| |||||||||||| ||| || ||||
Db   38382 TCTGCGAGGCATAGAGCCATTCCGTAGGGACATCTCGGTGCTTCTGAGGACGCGACCGA 38441

Qy   118 G--GGGATCTGATTA---TTGAGGTGTGGAAGGAATAAATAATCAGTCCACAAATAAAC 171
            | ||||||||||| |||||||||||||||||| ||||||||||||| |||||||
Db   38442 GCAGGGATCCGATGACGACTGGAGATGTTGAAGGAATAAAT-ACCAGTCCACAAATAAAC 38500

Qy   172 AAACTGT--CCGGGATTCCTAGAGGGAAGGAG 201
            |||||| |||||||||||||||||||||||
Db   38501 AAACTGTCCCGGGATTCCTAGAGGGAAGGAG 38532

RESULT 4
AC136621/c
LOCUS   AC136621   186021 bp   DNA   linear   PRI 31-OCT-2003
DEFINITION   Homo sapiens chromosome 16 clone RP11-481F24, complete sequence.
ACCESSION   AC136621
VERSION   AC136621.2   GI:38093724
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1   (bases 1 to 186021)
AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
        Alamos National Laboratory.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2   (bases 1 to 186021)

AUTHORS   DOE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Submitted (06-NOV-2002) Production Sequencing Facility, DOE Joint
        Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3   (bases 1 to 186021)
AUTHORS   Stanford Human Genome Center and Los Alamos National Laboratory.
CONSRTM   DOE Joint Genome Institute
TITLE   Direct Submission
JOURNAL   Submitted (31-OCT-2003) DOE Joint Genome Institute, 2800 Mitchell
        Drive, Walnut Creek, CA 94598, USA
COMMENT   On Oct 31, 2003 this sequence version replaced gi:24635938.
        Draft Sequence Produced by DOE Joint Genome Institute
        www.jgi.doe.gov
        Finishing Completed at Stanford Human Genome Center and Los Alamos
        National Laboratory
        www-shgc.stanford.edu
        Quality: Phrap Quality >=40 100% of Sequence;
        Estimated Total Number of Errors is 0.
FEATURES   Location/Qualifiers
source   1..186021
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="16"
        /clone="RP11-481F24"
ORIGIN

Query Match   18.9%;   Score 38;   DB 9;   Length 186021;
Best Local Similarity   60.8%;   Pred. No. 0.49;
Matches 62;   Conservative 0;   Mismatches 40;   Indels 0;   Gaps 0;

Qy   72 AGCCTTGTTAAAGGGGTGTCTCCACTCCTTCAGGGAAAGTGGGAAAAGGGGATCTGATTAT 131
            |||| ||||||||| ||||||| |||||||||||| ||||||| |||| ||||||||
Db   2791 AGATTGTTCAAGGAAGGTCCTCATGCTTAAGGAAAATGATGAAAATGGAACTACATAA 2732

Qy   132 TGAGGTGTGGAAGGAATAAATAATATCAGTCCACAAATAAACAA 173
            |||||||||||||||| |||| ||||||||||| |||||||
Db   2731 AGAAGAGTGCTAGAAACAGTAAATATGTAGATAAATATAAAA 2690

RESULT 5
AC026802
LOCUS   AC026802   188923 bp   DNA   linear   PRI 25-OCT-2003
DEFINITION   Homo sapiens chromosome 16 clone CTD-2545G24, complete sequence.
ACCESSION   AC026802
VERSION   AC026802.9   GI:37991843
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
        Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
        Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1   (bases 1 to 188923)
AUTHORS   DOE Joint Genome Institute, Stanford Human Genome Center and Los
        Alamos National Laboratory.
TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2   (bases 1 to 188923)
AUTHORS   DOE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Submitted (24-MAR-2000) Production Sequencing Facility, DOE Joint
        Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   3   (bases 1 to 188923)
AUTHORS   DOE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Submitted (08-JUL-2002) Production Sequencing Facility, DOE Joint
        Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   4   (bases 1 to 188923)
AUTHORS   DOE Joint Genome Institute.
TITLE   Direct Submission
JOURNAL   Submitted (17-DEC-2002) Production Sequencing Facility, DOE Joint
        Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE   5   (bases 1 to 188923)
AUTHORS   Stanford Human Genome Center and Los Alamos National Laboratory.

CONSRTM   DOE Joint Genome Institute
TITLE   Direct Submission
JOURNAL   Submitted (25-OCT-2003) DOE Joint Genome Institute, 2800 Mitchell
           Drive, Walnut Creek, CA 94598, USA
COMMENT   On Oct 25, 2003 this sequence version replaced gi:27151362.
           Draft Sequence Produced by DOE Joint Genome Institute
           www.jgi.doe.gov
           Finishing Completed at Stanford Human Genome Center and Los Alamos
           National Laboratory
           www-shgc.stanford.edu
           Quality: Phrap Quality >=40 100% of Sequence;
           Estimated Total Number of Errors is 0.
FEATURES   Location/Qualifiers
     source   1..188923
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="16"
                     /clone="CTD-2545G24"
ORIGIN

Query Match   18.9%;   Score 38; DB 9; Length 188923;
Best Local Similarity   60.8%;   Pred. No. 0.49;
Matches   62; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Qy   72 AGCCTTGTTAAAGGGGGTGTCTCACTCCTTCAGGGAAGTGGGAAAAGGGGATCTGATTAT 131
           || |||||| ||| ||||| || || ||||| ||| || |||||| ||||| ||
Db 145440 AGAATTGTTCAGGAGGTCCTCATGCTTAAGGAAAATGATGAAAAGTGGAACTATATAA 145499

Qy   132 TGAGGTGTGGAAGGAATAAATAATCAGTCCACAAATAAACAA 173
           | ||| | |||| || | || | | ||| |||| ||||
Db 145500 AGAAGAGTGCTAGAAACAGTAAATATGTAGATAAATATAAAA 145541

RESULT 6
AC084381/c
LOCUS   AC084381   129727 bp   DNA   linear   PRI 09-JAN-2002
DEFINITION Homo sapiens BAC clone RP11-20K20 from 7, complete sequence.
ACCESSION   AC084381
VERSION   AC084381.6 GI:15618716
KEYWORDS   HTG.
SOURCE   Homo sapiens (human)
  ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 129727)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE   Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   MEDLINE   99063792
   PUBMED   9847074
REFERENCE   2 (bases 1 to 129727)
  AUTHORS   Shah,N., Kozlowicz,A. and Elliott,G.
  TITLE   The sequence of Homo sapiens BAC clone RP11-20K20
  JOURNAL   Unpublished (2001)
REFERENCE   3 (bases 1 to 129727)
  AUTHORS   Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (30-OCT-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4 (bases 1 to 129727)
  AUTHORS   Waterston,R.H.
  TITLE   Direct Submission
  JOURNAL   Submitted (18-SEP-2001) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5 (bases 1 to 129727)
  AUTHORS   Waterston,R.
  TITLE   Direct Submission
  JOURNAL   Submitted (09-JAN-2002) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT   On Sep 18, 2001 this sequence version replaced gi:15144331.

           ---------- Genome Center
              Center: Washington University Genome Sequencing Center
              Center code: WUGSC
              Web site: http://genome.wustl.edu/gsc
              Contact: sapiens@watson.wustl.edu
           ---------- Summary Statistics
              Center project name: H_NH0020K20
           ----------.

           NOTICE: This sequence may not represent the entire insert of this
           clone. It may be shorter because we only sequence overlapping
           clone sections once, or longer because we provide a small overlap
           between neighboring data submissions.

           This sequence was finished as follows unless otherwise noted:
           all regions were double stranded, sequenced with an alternate
           chemistry, or covered by high quality data (i.e., phred quality >=
           30); an attempt was made to resolve all sequencing problems, such
           as compressions and repeats; all regions were covered by sequence
           from more than one subclone; and the assembly was confirmed by
           restriction digest.

           MAPPING INFORMATION:
           The sequence of this clone was established as part of a mapping and
           sequencing collaboration between the NHGRI Chromosome 7 Mapping
           Project (Eric D. Green, Director), John D. McPherson in the
           Department of Genetics (Washington University), and the Washington
           University Genome Sequencing Center. For additional information
           about the map position of this sequence, see
           http://www.nhgri.nih.gov/DIR/GTB/CHR7, send
           mailto:egreen@nhgri.nih.gov , or see http://genome.wustl.edu/gsc

           SOURCE INFORMATION:
           The RPCI-11 human BAC library was made from the blood of one male
           donor, as described by Osoegawa,K., Woon,P.Y., Zhao,B., Frengen,E.,
           Tateno,M., Catanese,J.J. and de Jong,P.J. (1998) An improved
           approach for construction of bacterial artificial chromosome
           libraries. Genomics 51:1-8. The clone may be obtained either from
           Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
           and coworkers at the Roswell Park Cancer Institute
           (http://bacpac.med.buffalo.edu)
           VECTOR: pBACe3.6
           NEIGHBORING SEQUENCE INFORMATION:
           The clone sequenced to the left is CTB-67M9, 2000 bp overlap; the
           clone sequenced to the right is RP11-115N4. Actual start of this
           clone is at base position 29168 of CTB-67M9; actual end is at base
           position 129727 of RP11-20K20.

           Polymorphisms exist between RP11-20K20 and CTB-67M9. Data from
           AC010764 was used to finish this clone, AC084381.
FEATURES   Location/Qualifiers
     source   1..129727
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="7"
                     /map="7"
                     /clone="RP11-20K20"
                     /clone_lib="RPCI-11"
     repeat_region   1..88
                     /rpt_family="L1"
     repeat_region   116..295
                     /rpt_family="(TCCCC)n"
     repeat_region   309..485
                     /rpt_family="L1"
     repeat_region   1519..2146
                     /rpt_family="ERV1"
     repeat_region   2530..2552
                     /rpt_family="(TGGA)n"
     repeat_region   3843..4010
                     /rpt_family="MIR"
     repeat_region   4110..4398
                     /rpt_family="MaLR"

```
repeat_region   4467..4775
                /rpt_family="Alu"
repeat_region   5480..5556
                /rpt_family="(TA)n"
repeat_region   6384..6452
                /rpt_family="MIR"
repeat_region   7703..7784
                /rpt_family="(TA)n"
repeat_region   10148..10267
                /rpt_family="MER1_type"
repeat_region   10261..10357
                /rpt_family="MER1_type"
repeat_region   11105..11406
                /rpt_family="Alu"
repeat_region   13246..13330
                /rpt_family="AT_rich"
misc_feature    15570..16155
                /note="similar to Rattus norvegicus EST BF413169
                (NID:g11401158)"
repeat_region   17320..17397
                /rpt_family="L2"
repeat_region   18230..18657
                /rpt_family="L1"
repeat_region   18658..18946
                /rpt_family="Alu"
repeat_region   18947..19698
                /rpt_family="L1"
repeat_region   19640..19682
                /rpt_family="AT_rich"
repeat_region   19773..19800
                /rpt_family="AT_rich"
misc_feature    21145..21594
                /note="match to EST BE925812 (NID:g10451888)"
misc_feature    21149..21424
                /note="similar to Mus musculus EST BF472109
                (NID:g15541292)"
misc_feature    21149..21420
                /note="similar to Homo sapiens EST BF352373
                (NID:g13311447)"
misc_feature    21149..21379
                /note="match to EST BE560547 (NID:g9804267)"
misc_feature    21193..21729
                /note="match to EST BE169347 (NID:g8632068)"
misc_feature    21361..21728
                /note="similar to Mus musculus EST BB850519
                (NID:g10308858) uw19b05.y1"
misc_feature    21181..21800
                /note="match to EST AW501161 (NID:g7114430)"
misc_feature    21833..22775
                /note="match to EST AL535374 (NID:g12798867)"
misc_feature    22190..22851
                /note="match to EST BG260591 (NID:g12770407)"
misc_feature    22343..22770
                /note="match to EST AI204989 (NID:g3758051) an03b07.x1"
misc_feature    22580..22808
                /note="match to EST BE765363 (NID:g10195287)"
misc_feature    22814..23472
                /note="match to EST BG573517 (NID:g13581170)"
misc_feature    23007..23672
                /note="similar to Homo sapiens EST BE545373
                (NID:g9774018)"
misc_feature    23049..23259
                /note="match to EST Z44576 (NID:g573722)"
repeat_region   23381..23450
                /rpt_family="CT-rich"
misc_feature    23491..24123
                /note="match to EST AL120642 (NID:g5926541)"
misc_feature    23523..23995
                /note="similar to Bos taurus EST AV663778 (NID:g9922808)"
misc_feature    23594..24395
                /note="match to EST AL565894 (NID:g12917721)"
misc_feature    23801..24550
                /note="similar to Rattus norvegicus EST BG671672

misc_feature    (NID:g13893771)"
                23815..24249
                /note="match to EST BE007039 (NID:g8267272)"
misc_feature    23993..24269
                /note="similar to Mus musculus EST BB192528
                (NID:g9851147)"
misc_feature    24000..24577
                /note="match to EST BI253803 (NID:g14805587)"
misc_feature    24018..24197
                /note="match to EST AA465707 (NID:g2191229) aa32e03.s1"
misc_feature    24021..24275
                /note="similar to Mus musculus EST BE955751
                (NID:g10599608)"
misc_feature    24025..24576
                /note="match to EST AA883846 (NID:g2993376) am21h10.s1"
misc_feature    24121..24572
                /note="similar to Mus musculus EST BG797711


Query Match            18.3%;  Score 36.8;  DB 9;  Length 129727;
Best Local Similarity  66.2%;  Pred. No. 1.2;
Matches  53;  Conservative  0;  Mismatches  27;  Indels  0;  Gaps  0;

Qy     21 CCTCTTCCCTGCCCAACCAGACCAGCAACTTCCAAAATTCTGCCCGTGTTTAGCCTTGTT 80
          ||||||||||| || ||  || || | ||  |  |||||| || ||  |    ||| |||
Db  93848 CCTCTTCCCTGGCAACAAATCAGCACCTCTGAAAATGAGGCCAGAAATCCTCATTGTT 93789

Qy     81 AAAGGGGTGTCTCACTCCTT 100
          || | | ||| ||   |  |
Db  93788 AACGTGCTCTCTGAGTGATT 93769


RESULT 7
AC131549
LOCUS       AC131549               241394 bp    DNA     linear   HTG 13-NOV-2002
DEFINITION  Rattus norvegicus clone CH230-11F15, *** SEQUENCING IN PROGRESS
            ***, 2 unordered pieces.
ACCESSION   AC131549
VERSION     AC131549.3  GI:24941082
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 241394)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,J., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,J., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
            Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
            Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
            Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
            Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guevara,W., Guevara,W.,
            Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
            Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
            Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
            Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
            Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
            Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
            Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
            Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
            Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
            Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
            Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
            Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
```

```
*   be preserved.
*          1    239695: contig of 239695 bp in length
*     239696    239795: gap of unknown length
*     239796    241394: contig of 1599 bp in length.
FEATURES             Location/Qualifiers
     source          1..241394
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-11F15"
     misc_feature    1..2347
                     /note="wgs_contig"
     misc_feature    31491..32938
                     /note="wgs_contig"
     misc_feature    35624..36681
                     /note="wgs_contig"
     misc_feature    36732..38223
                     /note="wgs_contig"
     misc_feature    102229..106221
                     /note="wgs_contig"
     misc_feature    131470..133327
                     /note="wgs_contig"
     misc_feature    215653..217923
                     /note="wgs_contig"
ORIGIN

Query Match     18.3%;   Score 36.8;   DB 2;   Length 241394;
Best Local Similarity  39.8%;   Pred. No. 1.2;
Matches  74;   Conservative  0;  Mismatches 112;  Indels  0;  Gaps  0;

Qy      3  AAATGTTCTCCCTGAGCCCCTTCCCTGCCCAACCAGACCAGCAACTTCCAAAATTCTG 62
           |||  ||| |||||| ||| |||| |||||||||||||||| ||||||||||||||||
Db  49007  AGATGCTATCCCTTAACCCTCTTTCCTGAACAACCCTCCCAAAAGCATCCATAGTTGTT 49066

Qy     63  CCCGTGTTAGCCTTGTTAAAGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAAGGGGA 122
           |  ||||||| ||| | | |||  |||| ||| |  | |  |||||  ||| | ||||
Db  49067  AAGGTACATTGCCTTCATACACTGTATGACTTGTGTCTAAGTTCTCTAGTTAATAAAGTACA 49126

Qy    123  TCTGATTATTGAGGTGTGGAAGGAATAAAATAATCAGTCCACAAATAAACAAACTGTCCGG 182
           | | || || | ||| ||  |||  |  ||||  |
Db  49127  ACTGTTAAAAATGGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 49186

Qy    183  GATTCC 188
           ||||
Db  49187  NNNNCC 49192

RESULT 8
AC119381
LOCUS       AC119381          261236 bp    DNA    linear   HTG 08-OCT-2002
DEFINITION  Rattus norvegicus clone CH230-161P20, *** SEQUENCING IN PROGRESS
            ***.
ACCESSION   AC119381
VERSION     AC119381.4  GI:22856286
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE      Rattus norvegicus (Norway rat)
  ORGANISM  Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 261236)
  AUTHORS   Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Alder,J.,
            Allen,C., Allen,H., Alebrooks,S., Amin,A., Anguiano,D.,
            Anyalebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
            Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
            Biswalo,K., Blair,D., Blankenburg,K., Blyth,P., Brown,M.,
            Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
            Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
            Chacko,J., Chavez,D., Chen,G., Chen,K., Chen,Y., Chen,Z., Chu,J.,
            Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
            Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
            Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
            Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
```

```
            Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
            Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
            Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S., Norris,S.,
            Nwaokelemeh,O., Okwuonu,G., Olanrpunsagoon.A., Pal,S., Parks,K.,
            Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
            Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
            Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
            Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
            Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose.R., Ruiz,S.J.,
            Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
            Shetty,J., Shvartsbeyn,A., Sisson.I., Sitter,C.D., Smajs,D.,
            Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
            Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
            Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
            Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,B., Wang,J.,
            Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
            Williams,G., Willson,R., Wieczyk,R., Wooden,H., Worley,K.,
            Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
            Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
            Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
            Weinstock,G. and Gibbs,R.A.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 241394)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-AUG-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 241394)
  AUTHORS   Rat Genome Sequencing Consortium.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-NOV-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Nov 13, 2002 this sequence version replaced gi:22855669.
            The sequence in this assembly is a combination of BAC based reads
            and whole genome shotgun sequencing reads assembled using Atlas
            (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
            in the feature table below represents a scaffold in the Atlas
            assembly (a 'contig-scaffold'). Within each contig-scaffold,
            individual sequence contigs are ordered and oriented, and separated
            by sized gaps filled with Ns to the estimated size. The sequence
            may extend beyond the ends of the clone and there may be sequence
            contigs within a contig-scaffold that consist entirely of whole
            genome shotgun sequence reads. Both end sequences and whole genome
            shotgun sequence only contigs will be indicated in the feature
            table.
            ----------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
            ----------- Project Information
            Center project name: GEBC
            Center clone name: CH230-11F15
            ----------- Summary Statistics
            Assembly program: Phrap; version 0.990329
            Consensus quality: 203660 bases at least Q40
            Consensus quality: 209296 bases at least Q30
            Consensus quality: 212200 bases at least Q20
            Estimated insert size: 215943; sum-of-contigs estimation
            Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            *   consists of 2 contigs. The true order of the pieces
            *   is not known and their order in this sequence record is
            *   arbitrary. Gaps between the contigs are represented as
            *   runs of Ns, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            *   as soon as it is available and the accession number will
```

Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,
Fraser,C.M., Gabisi,A., Ganta,S., Garcia,A., Garner,T., Garza,M.,
Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
Gunaratne,P., Haaland,W., Hamil,C., Hamilton,K.,
Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hogues,M.,
Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,B.,
Milosavljevic,A., Miner,G., Minja,B., Montemayor,J., Moore,S.,
Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
Nwaokelemeh,O., Okwuonu,G., Olarmpunsagoon,A., Pal,S., Parks,K.,
Pasternak,S., Paul,H., Perez,A., Perez,L., Pfannkoch,C.,
Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
Steinle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
Taylor,R., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
Valas,R., Vera,V., Villasana,D., Waldron,L., Walker,D., Wang,J.,
Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
Niederhausern,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.,
Weinstock,G. and Gibbs,R.A.

TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 261236)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (26-APR-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
REFERENCE 3 (bases 1 to 261236)
AUTHORS   Rat Genome Sequencing Consortium.
TITLE   Direct Submission
JOURNAL   Submitted (08-OCT-2002) Human Genome Sequencing Center, Department
          of Molecular and Human Genetics, Baylor College of Medicine, One
          Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Sep 14, 2002 this sequence version replaced gi:21903180.
          The sequence in this assembly is a combination of BAC based reads
          and whole genome shotgun sequencing reads assembled using Atlas
          (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
          in the feature table below represents a scaffold in the Atlas
          assembly (a 'contig-scaffold'). Within each contig-scaffold,
          individual sequence contigs are ordered and oriented, and separated
          by sized gaps filled with Ns to the estimated size. The sequence
          may extend beyond the ends of the clone and there may be sequence
          contigs within a contig-scaffold that consist entirely of whole
          genome shotgun sequence reads. Both end sequences and whole genome
          shotgun sequence only contigs will be indicated in the feature
          table.
          --------------- Genome Center
            Center: Baylor College of Medicine
            Center code: BCM
            Web site: http://www.hgsc.bcm.tmc.edu/
            Contact: hgsc-help@bcm.tmc.edu
          --------------- Project Information
            Center project name: GWHU
            Center clone name: CH230-161P20
          --------------- Summary Statistics

Assembly program: Phrap; version 0.990329
            Consensus quality: 235487 bases at least Q40
            Consensus quality: 238890 bases at least Q30
            Consensus quality: 240652 bases at least Q20
            Estimated insert size: 261654; sum-of-contigs estimation
            Quality coverage: 4x in Q20 bases; sum-of-contigs estimation
          ----
          * NOTE: Estimated insert size may differ from sequence length
          *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)
          * NOTE: This sequence may represent more than one clone.
          * NOTE: This is a 'working draft' sequence. It currently
          *   consists of 1 contigs. Gaps between the contigs
          *   are represented as runs of N. The order of the pieces
          *   is believed to be correct as given, however the sizes
          *   of the gaps between them are based on estimates that have
          *   provided by the submitter.
          * This sequence will be replaced
          * by the finished sequence as soon as it is available and
          * the accession number will be preserved.
          *   1 261236: contig of 261236 bp in length.
FEATURES   Location/Qualifiers
     source   1..261236
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-161P20"
     misc_feature   1..1435
                     /note="wgs_contig"
     misc_feature   1486..3278
                     /note="wgs_contig"
ORIGIN

Query Match
Best Local Similarity 18.3%;   Score 36.8;   DB 2;   Length 261236;
Matches   74; Conservative   0; Mismatches   62; Indels   0; Gaps   0;

Qy   3 AAATGTTCTCCTGAGCCCTCTTCCTGCCACCAGACCAGCAACTTCCAAAATTCTG 62
            ||||| |||||||| |||||| |||| ||| || | ||| ||||||||||||
Db 143400 AGATGCTATCCCTTAAACCCTCTTTCCTGAACACCACCCTCCCAAAAGCATCCATAGTTGTT 143459

Qy   63 CCCGTGTTTAGCCTTGTTAAAGGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAAGGGGA 122
            ||||| ||| ||| | || |||||| || | ||||||||||||| ||| |||||||||
Db 143460 AAGGTACATTGCCTTCATCACTGATGACTTGTGTCTAAGTTCGTTCAGTTAATAAGTACA 143519

Qy   123 TCTGATTATTGAGGTG 138
            || ||| |||||
Db 143520 ACTGTTTAAAAATGGTG 143535

RESULT 9
AC023374
LOCUS   AC023374   128680 bp   DNA   linear   HTG 27-APR-2000
DEFINITION Homo sapiens clone RP11-22F4, *** SEQUENCING IN PROGRESS ***, 74
            unordered pieces.
ACCESSION   AC023374
VERSION   AC023374.2 GI:7657735
KEYWORDS   HTG; HTGS_PHASE1.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 128680)
AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE   Homo sapiens chromosome, clone RP11-22F4
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 128680)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beda,F., Boguslavkiy,L.,
            Boukhgalter,B., Brown,A., Burkett,G., Campopiano,A., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Dodge,S., Domino,M., Doyle,M.,
            Fenestor,J., Ferreira,P., FitzHugh,W., Forrest,C., Gage,D.,
            Galagar,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,

```
           Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
           Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
           Klein,J., Landers,T., Largocque,K., Lehoczky,J., Levine,R.,
           Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N., McCarthy,M.,
           McEwan,P., McGurk,A., McKernan,K., McPheeters,R., Meldrim,J.,
           Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J., Naylor,J.,
           Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Olivar,T.M.,
           Peterson,K., Pierre,N., Pisani,C., Pollara,V., Raymond,C.,
           Riley,R., Rogov,P., Rothman,D., Roy,A., Santos,R., Schauer,S.,
           Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
           Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J., Tirrell,A.,
           Travers,M., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B.,
           Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zimmer,A. and
           Zody,M.
TITLE      Direct Submission
JOURNAL    Submitted (14-FEB-2000) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT    On Apr 27, 2000 this sequence version replaced gi:6970507.
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           ----------- Genome Center
           Center: Whitehead Institute/ MIT Center for Genome Research
           Center code: WIBR
           Web site: http://www-seq.wi.mit.edu
           Contact: sequence_submission@genome.wi.mit.edu
           --------- Project Information
           Center project name: L4113
           Center clone name: 22_F_4
           ---
         * NOTE: This is a 'working draft' sequence. It currently
         * consists of 74 contigs. The true order of the pieces
         * is not known and their order in this sequence record is
         * arbitrary. Gaps between the contigs are represented as
         * runs of N, but the exact sizes of the gaps are unknown.
         * This record will be updated with the finished sequence
         * as soon as it is available and the accession number will
         * be preserved.
         *        1:  1301: contig of 1301 bp in length
         *     1302:   140: gap of 100 bp
         *     1402:  2463: contig of 1062 bp in length
         *     2464:  2563: gap of 100 bp
         *     2564:  3612: contig of 1049 bp in length
         *     3613:  3712: gap of 100 bp
         *     3713:  4785: contig of 1073 bp in length
         *     4786:  4885: gap of 100 bp
         *     4886:  5996: contig of 1111 bp in length
         *     5997:  6096: gap of 100 bp
         *     6097:  7466: contig of 1370 bp in length
         *     7467:  7566: gap of 100 bp
         *     7567:  8610: contig of 1044 bp in length
         *     8611:  8710: gap of 100 bp
         *     8711:  9913: contig of 1203 bp in length
         *     9914: 10013: gap of 100 bp
         *    10014: 11040: contig of 1027 bp in length
         *    11041: 11140: gap of 100 bp
         *    11141: 12424: contig of 1284 bp in length
         *    12425: 12524: gap of 100 bp
         *    12525: 13822: contig of 1298 bp in length
         *    13823: 13922: gap of 100 bp
         *    13923: 14934: contig of 1012 bp in length
         *    14935: 15034: gap of 100 bp
         *    15035: 16184: contig of 1150 bp in length
         *    16185: 16284: gap of 100 bp
         *    16285: 17291: contig of 1007 bp in length
         *    17292: 17391: gap of 100 bp
         *    17392: 18472: contig of 1081 bp in length
         *    18473: 18572: gap of 100 bp
         *    18573: 19722: contig of 1149 bp in length
         *    19722: 19821: gap of 100 bp
         *    19822: 21058: contig of 1237 bp in length
         *    21058: 21158: gap of 100 bp
         *    21159: 22325: contig of 1167 bp in length
         *    22326: 22425: gap of 100 bp
         *    22426: 22543: contig of 118 bp in length
         *    22544: 22643: gap of 100 bp
         *    22644: 24242: contig of 1599 bp in length
         *    24243: 24342: gap of 100 bp
         *    24343: 25953: contig of 1611 bp in length
         *    25954: 26053: gap of 100 bp
         *    26054: 27302: contig of 1249 bp in length
         *    27303: 27402: gap of 100 bp
         *    27403: 28790: contig of 1388 bp in length
         *    28791: 28890: gap of 100 bp
         *    28891: 30136: contig of 1246 bp in length
         *    30137: 30236: gap of 100 bp
         *    30237: 31616: contig of 1380 bp in length
         *    31617: 31716: gap of 100 bp
         *    31717: 33518: contig of 1802 bp in length
         *    33519: 33618: gap of 100 bp
         *    33619: 34937: contig of 1319 bp in length
         *    34938: 35037: gap of 100 bp
         *    35038: 36372: contig of 1335 bp in length
         *    36373: 36472: gap of 100 bp
         *    36473: 38358: contig of 1886 bp in length
         *    38359: 38458: gap of 100 bp
         *    38459: 39477: contig of 1019 bp in length
         *    39478: 39577: gap of 100 bp
         *    39578: 40629: contig of 1052 bp in length
         *    40630: 40729: gap of 100 bp
         *    40730: 41995: contig of 1266 bp in length
         *    41996: 42095: gap of 100 bp
         *    42096: 43379: contig of 1284 bp in length
         *    43380: 43479: gap of 100 bp
         *    43480: 44865: contig of 1386 bp in length
         *    44866: 44965: gap of 100 bp
         *    44966: 46815: contig of 1850 bp in length
         *    46816: 46915: gap of 100 bp
         *    46916: 48514: contig of 1599 bp in length
         *    48515: 48614: gap of 100 bp
         *    48615: 50670: contig of 2056 bp in length
         *    50671: 50770: gap of 100 bp
         *    50771: 51821: contig of 1051 bp in length
         *    51822: 51921: gap of 100 bp
         *    51922: 54151: contig of 2230 bp in length
         *    54152: 54251: gap of 100 bp
         *    54252: 55643: contig of 1392 bp in length
         *    55644: 55743: gap of 100 bp
         *    55744: 57332: contig of 1589 bp in length
         *    57333: 57432: gap of 100 bp
         *    57433: 58921: contig of 1489 bp in length
         *    58922: 59021: gap of 100 bp
         *    59022: 60563: contig of 1542 bp in length
         *    60564: 60663: gap of 100 bp
         *    60664: 62029: contig of 1366 bp in length
         *    62030: 62129: gap of 100 bp
         *    62130: 63463: contig of 1334 bp in length
         *    63464: 63563: gap of 100 bp
         *    63564: 65051: contig of 1488 bp in length
         *    65052: 65151: gap of 100 bp
         *    65152: 66808: contig of 1657 bp in length
         *    66809: 66908: gap of 100 bp
         *    66909: 68875: contig of 1967 bp in length
         *    68876: 68975: gap of 100 bp
         *    68976: 70346: contig of 1371 bp in length
         *    70347: 70446: gap of 100 bp
         *    70447: 72459: contig of 2013 bp in length
         *    72460: 72559: gap of 100 bp
         *    72560: 74054: contig of 1495 bp in length
         *    74055: 74154: gap of 100 bp
         *    74155: 75973: contig of 1819 bp in length
         *    75974: 76073: gap of 100 bp
         *    76074: 77770: contig of 1697 bp in length
         *    77771: 77870: gap of 100 bp
         *    77871: 79793: contig of 1923 bp in length
         *    79794: 79893: gap of 100 bp
```

```
  *    79994  82188: contig of 2295 bp in length
  *    81189  82288: gap of 100 bp
  *    82289  84708: contig of 2420 bp in length
  *    84709  84808: gap of 100 bp
  *    84809  86868: contig of 2060 bp in length
  *    86869  86968: gap of 100 bp
  *    86969  89135: contig of 2167 bp in length
  *    89136  89235: gap of 100 bp
  *    89236  91440: contig of 2205 bp in length
  *    91441  91540: gap of 100 bp
  *    91541  93441: contig of 1901 bp in length
  *    93442  93541: gap of 100 bp
  *    93542  95044: contig of 1503 bp in length
  *    95045  95144: gap of 100 bp
  *    95145  96925: contig of 1781 bp in length
  *    96926  97025: gap of 100 bp
  *    97026  98950: contig of 1925 bp in length
  *    98951  99050: gap of 100 bp
  *    99051 101096: contig of 2046 bp in length
  *   101097 101196: gap of 100 bp
  *   101197 101901: contig of 705 bp in length
  *   101902 102001: gap of 100 bp
  *   102002 104575: contig of 2574 bp in length
  *   104576 104675: gap of 100 bp
  *   104676 107733: contig of 3058 bp in length
  *   107734 107833: gap of 100 bp
  *   107834 110783: contig of 2950 bp in length
  *   110784 110883: gap of 100 bp
  *   110884 112896: contig of 2013 bp in length
  *   112897 112996: gap of 100 bp
  *   112997 115653: contig of 2657 bp in length
  *   115654 115753: gap of 100 bp

Query Match      18.1%;   Score 36.4;   DB 2;   Length 128680;
Best Local Similarity   54.5%;   Pred. No. 1.6;
Matches  73;  Conservative  0;  Mismatches  61;  Indels  0;  Gaps  0;

Qy      65 CGTGTTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAAGGGGATC 124
           ||||  ||| |||| ||| ||  ||  |||| || ||||||||||| |||| ||||||||
Db  116191 CGTAGAGAGCTAGGTCATATCGAGACTCACTGAATCAAAGTATTTGGGAAAAGGTTATC 116250

Qy     125 TGATTATTGAGGTGTGGAAGGAATAAATAATCAGTCCACAAATAAACAAACTGTCCGGGA 184
           || ||||| ||| || ||| |||| |||| ||| ||  |||| ||| | | ||  ||||
Db  116251 TGAAAATGGATATATGGGGAGCAAATATTTAATATTCCCGTAATTATTATGTACCTTAAAA 116310

Qy     185 TTCCTAGAGGGAAG 198
           || ||| ||| ||
Db  116311 TTCGAAGATTGATG 116324


RESULT 10
AC009247/c
LOCUS       AC009247           146376 bp   DNA     linear    PRI 14-DEC-2001
DEFINITION  Homo sapiens 3q27  BAC RP11-410F19 (Roswell Park Cancer Institute
            Human BAC Library) complete sequence.
ACCESSION   AC009247
VERSION     AC009247.13  GI:17737024
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 146376)
  AUTHORS   Muzny,D.M., Adams,C., Bailey,M., Barbaria,J., Blankenburg,K.,
            Bodcta,B., Bouck,J., Bowie,S., Brooks,A., Buhay,C., Bunac,C.,
            Burkett,C., Burrows,J., Carter,M., Chacko,J., Chen,Z., Cox,C.,
            David,R., Delgado,O., Deshazo,D., Ding,Y., Domah-Rashid,N.,
            Dugan-Rocha,S., Durbin,K.J., Fernandez,C., Ferraguto,D.,
            Forcum-Tansey,J., Frantz,P., Ganesh,R., Garcia,D.K., Gorrell,J.H.,
            Gorrell,L.L., Guevara,W., Harris,K., He,X., Hernandez,J.,
            Hodgson,A., Hogues,M., Holloway,C., Hosak,H., Jackson,L.E.,
            Jackson,L., Jia,Y., Jones,M., Kelly,S., Kondejewski,N., Kong,Y.,
            Kovar,C., Leal,B., Li,Z., Lichtarge,O., Liu,J., Liu,W., Logan,O.,
```

Lozado,R.J., Lu,J., Lucier,R., Martin,R., Martinez,C., McLeod,M.P.,
Mei,G., Moore,S., Moorish,T., Morgan,M., Morris,S., Nash,S.,
Nelson,A., Nguyen,R., Nguyen,N., Nguyen,S., Oswal,G., Parish,B.,
Paxton,S., Payton,B., Perez,L., Pu,L.L., Quiles,M., Reiter,D.,
Rives,M., Samuel,S., Say,J., Scherer,S., Shah,E., Shen,H.,
Simon,M., Sparks,A., Stamps,A., Sucgang,R., Tabor,P., Taylor,T.,
Vasquez,L., Vinson,R., Vo,Q., Wahbah,M., Watlington,S.,
Weinstock,L., Weinstock,I.R., Williamson,A., Worley,K., Wren,J.,
Wrensford,G., Yu,W., Zhou,X., Naylor,S.L., Nelson,D. and Gibbs,R.

```
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 146376)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-AUG-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 146376)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-DEC-1999) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4  (bases 1 to 146376)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAR-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5  (bases 1 to 146376)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-DEC-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
            On Dec 14, 2001 this sequence version replaced gi:13489132.
COMMENT     INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
            STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
            Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
            Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:
            http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.
```

```
FEATURES       QUALSTAT·REPORT.
               Location/Qualifiers
source         1..146376
               /organism="Homo sapiens"
               /mol_type="genomic DNA"
               /db_xref="taxon:9606"
               /clone="RP11-410F19"
repeat_region  789..825
               /rpt_family="AT_rich"
misc_feature   1371..1711
               /note="Region: Similar to Hs#S1074034 oz46g11.x1 Homo
               sapiens cDNA, AI079997"
repeat_region  1737..1781
               /rpt_family="(CA)n"
repeat_region  3500..3520
               /rpt_family="AT_rich"
misc_feature   join(4659..4721,6288..6513)
               /note="Region: Similar to Hs#S1609703 DKFZp762L097_r1 Homo
               sapiens cDNA, AL121466"
repeat_region  5272..5325
               /rpt_family="GC_rich"
repeat_region  6178..6278
               /rpt_family="GC_rich"
repeat_region  6495..6637
               /rpt_family="C-rich"
repeat_region  6695..6718
               /rpt_family="GC_rich"
repeat_region  7005..7142
               /rpt_family="G-rich"
repeat_region  8361..8487
               /rpt_family="MIR"
repeat_region  9046..9067
               /rpt_family="(TTG)n"
repeat_region  complement(9153..9513)
               /rpt_family="THE1B"
repeat_region  complement(10270..10517)
               /rpt_family="AluJo"
repeat_region  complement(10736..11062)
               /rpt_family="MER1B"
repeat_region  11101..11155
               /rpt_family="MIR"
repeat_region  11362..11452
               /rpt_family="L1MB8"
repeat_region  complement(11696..11997)
               /rpt_family="MER5A"
repeat_region  12202..12827
               /rpt_family="L2"
repeat_region  12980..13267
               /rpt_family="HAL1"
repeat_region  13473..13671
               /rpt_family="AluJo"
repeat_region  complement(14040..14312)
               /rpt_family="L2"
repeat_region  complement(14495..14807)
               /rpt_family="AluJb"
repeat_region  complement(14897..15005)
               /rpt_family="AluSp"
repeat_region  complement(15090..15172)
               /rpt_family="L1MD3"
repeat_region  complement(15182..15221)
               /rpt_family="L1MC4"
repeat_region  15225..15533
               /rpt_family="L1MC4"
repeat_region  complement(15534..15650)
               /rpt_family="AluSq"
repeat_region  15651..15954
               /rpt_family="L1MC4"
repeat_region  complement(15955..16485)
               /rpt_family="AluYa5"
repeat_region  complement(16586..16678)
               /rpt_family="L1MC4"
repeat_region  complement(16679..16966)
               /rpt_family="AluSq/x"
               /rpt_family="L1MC4"
repeat_region  17043..17351
               /rpt_family="AluSg1"
misc_feature   17320..17410
               /function="Low Coverage"
repeat_region  17376..17407
               /rpt_family="(TTTG)n"
repeat_region  complement(17408..17699)
               /rpt_family="AluSx"
repeat_region  complement(17700..17843)
               /rpt_family="L1MC4"
repeat_region  complement(17878..18184)
               /rpt_family="AluSg"
repeat_region  complement(18422..18733)
               /rpt_family="AluY"
repeat_region  complement(18919..20097)
               /rpt_family="L1ME"
repeat_region  20107..20275
               /rpt_family="L1PA13"
repeat_region  20276..20580
               /rpt_family="AluSx"
repeat_region  20581..20741
               /rpt_family="L1PA13"
repeat_region  20758..20865
               /rpt_family="HAL1"
repeat_region  20879..20913
               /rpt_family="(TTTTA)n"
repeat_region  complement(20914..21210)
               /rpt_family="AluSg"
repeat_region  complement(21253..21311)
               /rpt_family="LTR33"
repeat_region  complement(21334..21422)
               /rpt_family="MER2"
repeat_region  21423..21723
```

```
Query Match        18.1%;  Score 36.4;  DB 9;  Length 146376;
Best Local Similarity  66.7%;  Pred. No. 1.6;
Matches 52;  Conservative   0;  Mismatches  26;  Indels   0;  Gaps   0;

Qy      7 GTTCTCCCTGAGAGCCCTCTTCCCTGCCCAACCAGACCAGACCAGCAACTTCCAAAATTCTGCCCG 66
          |||||||||||||||||||||||| ||||||||||||||||||||| |||||||||||||||||||
Db  61915 GTTCTGCCTGAGGCCGTTTCTCTTCTCTCCACCCTCTCCTGAAACATGCTAATTTCTGACTG 61856

Qy     67 TGTTTAGCCTTGTTAAAG 84
          ||||  ||  ||||||||
Db  61855 TGGCTTCCATTGTCATAG 61838
```

```
RESULT 11
AC022832/c
LOCUS       AC022832          166236 bp     DNA     linear   PRI 11-DEC-2001
DEFINITION  Homo sapiens chromosome 8, clone RP11-145015, complete sequence.
ACCESSION   AC022832
VERSION     AC022832.13  GI:17488715
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 166236)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 8, clone RP11-145015
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 166236)
  AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
            Anderson,S., Baldwin,J., Barna,N., Beckerly,R., Beda,F.,
            Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
            Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cooke,P.,
            DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenestor,J.,
            Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
            Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
            Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
            Landers,T., Lehoczky,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
            Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
```

```
                                                          Center clone name: 145_O_15

FEATURES             Location/Qualifiers
source               1..166236
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="8"
                     /map="8"
                     /clone="RP11-145O15"
                     /clone_lib="RPCI-11 Human Male BAC"
repeat_region        1..625
                     /rpt_family="L1M2"
repeat_region        634..1756
                     /rpt_family="L1M2"
repeat_region        1948..1972
                     /rpt_family="AT_rich"
repeat_region        2462..2763
                     /rpt_family="AluSp"
repeat_region        3171..3345
                     /rpt_family="MER90"
repeat_region        complement(4558..4955)
                     /rpt_family="MSTB"
repeat_region        5371..5426
                     /rpt_family="(TA)n"
repeat_region        5807..6249
                     /rpt_family="MLTIC"
repeat_region        6858..7180
                     /rpt_family="MLT1H"
repeat_region        8186..8474
                     /rpt_family="L2"
repeat_region        8475..8499
                     /rpt_family="(TA)n"
repeat_region        8500..8560
                     /rpt_family="L2"
repeat_region        8561..8741
                     /rpt_family="MER5A"
repeat_region        8742..9382
                     /rpt_family="L2"
repeat_region        10313..10477
                     /rpt_family="L1M4b"
repeat_region        10478..10795
                     /rpt_family="AluSx"
repeat_region        10796..11161
                     /rpt_family="L1M4b"
repeat_region        complement(12590..12762)
                     /rpt_family="Charlie2"
repeat_region        complement(12769..13037)
                     /rpt_family="Charlie2b"
repeat_region        13036..13330
                     /rpt_family="AluSx"
repeat_region        complement(13331..14033)
                     /rpt_family="Charlie2b"
repeat_region        complement(15291..15432)
                     /rpt_family="Charlie2b"
repeat_region        16710..16730
                     /rpt_family="AT_rich"
repeat_region        complement(17047..17264)
                     /rpt_family="MSTB2"
repeat_region        17267..18244
                     /rpt_family="MER11A"
repeat_region        complement(18248..18478)
                     /rpt_family="MSTB2"
repeat_region        18700..18720
                     /rpt_family="AT_rich"
repeat_region        19546..19828
                     /rpt_family="L1ME1"
repeat_region        19844..20095
                     /rpt_family="L1ME2"
repeat_region        21943..22028
                     /rpt_family="MER65A"
repeat_region        complement(22160..22512)
                     /rpt_family="L1MEc"
```

TITLE   Direct Submission
JOURNAL   Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 166236)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
           Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
           Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
           Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
           Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
           Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
           Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
           Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
           Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
           Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
           MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
           McCarthy,M., Macdonald,P., McKernan,K., McPheeters,R., Meldrim,J.,
           Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
           Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
           Oliver,J., Peterson,K., Phunkharg,P., Pierre,N., Pollara,V.,
           Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
           Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
           Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
           Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
           Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
           Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
           Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (02-SEP-2001) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 4 (bases 1 to 166236)
AUTHORS   Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N.,
           Anderson,S., Barna,N., Bastien,V., Boguslavkiy,L., Boukhgalter,B.,
           Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
           Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
           Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
           Ferreira,P., FitzHugh,W., Gage,D., Galagan,J., Gardyna,S.,
           Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
           Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
           Jones,C., Kamat,A., Karatas,A., Kells,C., LaRocque,K.,
           Lamazares,R., Landers,T., Lehoczky,J., Levine,R., Liu,G.,
           MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C.,
           McCarthy,M., Macdonald,P., McKernan,K., McPheeters,R., Meldrim,J.,
           Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
           Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D.,
           Oliver,J., Peterson,K., Phunkharg,P., Pierre,N., Pollara,V.,
           Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
           Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R.,
           Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
           Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
           Topham,K., Travers,M., Travis,N., Trigilio,J., Vassiliev,H.,
           Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
           Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE   Direct Submission
JOURNAL   Submitted (11-DEC-2001) Whitehead Institute/MIT Center for Genome
           Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On Dec 11, 2001 this sequence version replaced gi:15421017.
           All repeats were identified using RepeatMasker:
           Smit, A.F.A. & Green, P. (1996-1997)
           http://ftp.genome.washington.edu/RM/RepeatMasker.html
           ---------- Genome Center
               Center: Whitehead Institute/ MIT Center for Genome Research
               Center code: WIBR
               Web site: http://www-seq.wi.mit.edu
               Contact: sequence_submissions@genome.wi.mit.edu
           ---------- Project Information
               Center project name: L6140

Also at top of references list above REFERENCE 3:
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Olivar,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.

```
repeat_region    complement(22793..23107)
                 /rpt_family="L1R33"
repeat_region    complement(23947..24097)
                 /rpt_family="L1MEc"
repeat_region    24392..24463
                 /rpt_family="MER5A"
repeat_region    complement(24860..25243)
                 /rpt_family="THE1C"
repeat_region    complement(26003..26306)
                 /rpt_family="AluJb"
repeat_region    27314..27455
                 /rpt_family="L1ME1"
repeat_region    27456..27742
                 /rpt_family="AluSq"
repeat_region    27746..27780
                 /rpt_family="(TAAA)n"
repeat_region    27781..27836
                 /rpt_family="L1ME1"
repeat_region    complement(27837..28860)
                 /rpt_family="Tigger3b"
repeat_region    28861..29160
                 /rpt_family="AluY"
repeat_region    complement(29161..29350)
                 /rpt_family="Tigger3b"
repeat_region    29351..29557
                 /rpt_family="L1ME1"
repeat_region    29458..29604
                 /rpt_family="L1MA10"
repeat_region    complement(30649..30674)
                 /rpt_family="MIR"


Query Match        18.1%;   Score 36.4;   DB 9;   Length 166236;
Best Local Similarity  54.5%;   Pred. No.1.6;
Matches  73;  Conservative  0;  Mismatches  61;  Indels  0;  Gaps  0;

Qy   65    CGTGTTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAAGTGGGAAAAGGGGATC  124
           ||  ||| ||| || ||| ||  | | ||| | |||  |||| ||  ||||||||  |||
Db   57908 CGTAGAGAGCTAGGTCATATCTGAGACTCACTGAATCAAAGTATTTCGGGAAAAGGTTATC  57849

Qy   125   TGATTATTGAGGTGTGGAAGGAAGAATAAAATAATCAGTCACAAATAAACAAACTGTCCGGGA  184
           ||| | |||| | ||  |||  | | ||| | ||  | |  |||||| |  ||    | ||
Db   57848 TGAAAATGGATATATGGGAGCAAATATTTAAATATTCCCGTAATTATTATGTACCTTAAAA  57789

Qy   185   TTCCTAGGAGGAAG  198
           ||| | |  | ||
Db   57788 TTCGAAGATTGATG  57775


RESULT 12
AC104476/c
LOCUS       AC104476      163579 bp    DNA     linear   HTG  21-SEP-2002
DEFINITION  Pan troglodytes clone RP43-139N13, WORKING DRAFT SEQUENCE, 3
            ordered pieces.
ACCESSION   AC104476
VERSION     AC104476.2  GI:23266279
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Pan troglodytes (chimpanzee)
  ORGANISM  Pan troglodytes
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
REFERENCE   1  (bases 1 to 163579)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Cariaga,K., Coleman,B., Dietrich,N.L., Granite,S., Guan,X.,
            Gupta,J., Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R.,
            Karlins,E., Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L.,
            Maduro,V.B., Margulies,E.H., Masiello,C., Maskeri,B.,
            Mastrian,S.D., McCloskey,J.C., McDowell,J., Pagurigan,C.,
            Pearson,R., Portnoy,M.E., Prasad,A., Reddix-Dugue,N.,
            Schueler,M.G., Sison,C., Stantripop,S., Thomas,J.W., Thomas,P.J.,
            Touchman,J.W., Vogt,J.L., Walker,M., Wetherby,K.D., Wiggins,L.,
            Young,A., Zhang,L.-H. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 163579)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-DEC-2001) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 163579)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-SEP-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Sep 21, 2002 this sequence version replaced gi:17530718.
            --------------- Genome Center

            Center: NIH Intramural Sequencing Center
            Center code: NISC
            Web site: http://www.nisc.nih.gov
            Contact: nisc_zoo@nhgri.nih.gov
            --------------- Project Information
            Center project name: cez
            Center clone name: 139N13
            ---------------
            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies. the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            --------------- Summary Statistics
            Sequencing vector: plasmid; n/a; 100% of reads
            Chemistry: Dye-terminator Big Dye; 100% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 163300 bases at least Q40
            Consensus quality: 163363 bases at least Q30
            Consensus quality: 163375 bases at least Q20
            Insert size: 117000; agarose-fp
            Insert size: 163379; sum-of-contigs
            Quality coverage: 12.68x in Q20 bases; agarose-fp
            Quality coverage: 9.08x in Q20 bases; sum-of-contigs
            ---------------
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submitter.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1      60466: contig of 60466 bp in length
            *   60467      60566: gap of unknown length
            *   60567      69329: contig of 8763 bp in length
            *   69330      69429: gap of unknown length
            *   69430     163579: contig of 94150 bp in length
            ---------------
FEATURES             Location/Qualifiers
     source          1..163579
                     /organism="Pan troglodytes"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9598"
                     /clone="RP43-139N13"
                     /clone_lib="RP43"
     misc_feature    1..60466
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:left"
     misc_feature    1..30537
                     /note="clone overlaps with GenBank Accession Number
                     AC104154 clone RP43-108N2 (center project name cey)"
```

```
     misc_feature     60567..69329
                      /note="assembly_fragment"
     misc_feature     69430..163579
                      /note="assembly_fragment"
                      clone_end:T7
                      vector_side:right"

ORIGIN

Query Match                17.5%;  Score 35.2;  DB 2;  Length 163579;
Best Local Similarity  65.0%;  Pred. No. 4;
Matches    52;  Conservative    0;  Mismatches    28;  Indels    0;  Gaps    0;

Qy       21 CCTCTTCCCTGCCCAACCAGACCAGCAACTTCCAAAATTCTGCCCGTGTTTAGCCTTGTT 80
            ||||| ||||||||| |  ||||||| |||  |||| ||||||  ||||| ||||||| |||||
Db   163237 CCTCTTCCCTGGCCAACCAAAATCAGCACCTCTGAAAATGAGGCCAGAAATCCTCATTGTT 163178

Qy       81 AAAGGGGTGTCTCACTCCTT 100
            | ||||||||||||| ||||
Db   163177 AACATGCTCTCTGAGTGATT 163158


RESULT 13
BX000462/c
LOCUS        BX000462           167671 bp    DNA    linear    VRT 30-AUG-2003
DEFINITION   Zebrafish DNA sequence from clone DKEY-114F6 in linkage group 1,
             complete sequence.
ACCESSION    BX000462
VERSION      BX000462.10  GI:34368570
KEYWORDS     HTG.
SOURCE       Danio rerio (zebrafish)
ORGANISM     Danio rerio
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
             Cypriniformes; Cyprinidae; Danio.
REFERENCE    1  (bases 1 to 167671)
AUTHORS      Howden,P.
TITLE        Direct Submission
JOURNAL      Submitted (30-AUG-2003) Wellcome Trust Sanger Institute, Hinxton,
             Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
             zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
             On Aug 30, 2003 this sequence version replaced gi:33414474.
COMMENT      -------------- Genome Center
             Center: Wellcome Trust Sanger Institute
             Center code: SC
             Web site: http://www.sanger.ac.uk
             Contact: zfish-help@sanger.ac.uk
             --------------
             During sequence assembly data is compared from overlapping clones.
             Where differences are found these are annotated as variations
             together with a note of the overlapping clone name. Note that the
             variation annotation may not be found in the overlapping clone
             corresponding to the overlapping clone, as we submit sequences with
             only a small overlap as described above.
             This sequence was finished as follows unless otherwise noted: all
             regions were either double-stranded or sequenced with an alternate
             chemistry or covered by high quality data (i.e., phred quality >=
             30); an attempt was made to resolve all sequencing problems, such
             as compressions and repeats; all regions were covered by at least
             one plasmid subclone or more than one M13 subclone; and the
             assembly was confirmed by restriction digest, except on the rare
             occasion of the clone being a YAC.
             The following abbreviations are used to associate primary accession
             numbers given in the feature table with their source databases:
             Em:, EMBL; Sw:, SWISSPROT; Tr:, TREMBL; Wp:, WORMPEP; Information
             on the WORMPEP database can be found at
             http://www.sanger.ac.uk/Projects/C_elegans/wormpep Clone-derived
             Zebrafish pUC subclones occasionally display inconsistency over the
             length of mononucleotide A/T runs and conserved TA repeats. Where
             this is found the longest good quality representation will be
             submitted.
             Repeat names beginning 'Dr' were identified by the Recon repeat
             discovery system (Zhirong Bao and Sean Eddy, submitted), and those
             beginning 'drr' were identified by Rick Waterman (Stephen Johnson
```

```
             lab, WashU).  For further information see
             http://www.sanger.ac.uk/Projects/D_rerio/fishmask.shtml DKEY-114F6
             is from a Zebrafish BAC library
             VECTOR: pIndigoBAC-5.
FEATURES             Location/Qualifiers
     source          1..167671
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-114F6"
                     /clone_lib="DanioKey"

ORIGIN

Query Match                17.5%;  Score 35.2;  DB 5;  Length 167671;
Best Local Similarity  53.7%;  Pred. No. 4;
Matches    73;  Conservative    0;  Mismatches    63;  Indels    0;  Gaps    0;

Qy       56 AATTCTGCCCGTGTTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAA 115
            ||||||| |||||||||||||||||| |||||||| || ||| | | | ||||||||||||
Db    15112 AATTATGCTCGCTGATTAGCCTAATTAAAGCTTAGCTTATTCCACTTAAGGATTAGGGGCTA 15053

Qy      116 AAGGGGATCTGATTATTGAGGTGTGGAAGGAAATAAATAATCAGTCCACAAATAAACAAAC 175
            ||  | ||  |||   ||||||||||| | ||||||||||||||| ||||||||||||
Db    15052 AAGTAAGCCTGATTATAATGTGAGTGTGTTCAGAAATCATAACTGTACAATTCCCTGGTC 14993

Qy      176 TGTCCGGATTCCTAG 191
            || ||| || | ||
Db    14992 TGCCCATGAGACAGAG 14977


RESULT 14
AC079310
LOCUS        AC079310           180971 bp    DNA    linear    PRI 27-APR-2002
DEFINITION   Homo sapiens 12 BAC RP11-850F7 (Roswell Park Cancer Institute Human
             BAC Library) complete sequence.
ACCESSION    AC079310
VERSION      AC079310.20  GI:12000443
KEYWORDS     HTG.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 180971)
AUTHORS      Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
             Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbaria,J.,
             Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
             Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,A., Buhay,C.,
             Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
             Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
             Chen,Z., Chiu,D., Chowdhry,I., Christopoulos,C., Cleveland,C.D.,
             Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
             Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
             Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
             Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
             Elhaj,C., Emerling,S., Escotto,M., Falls,T., Ferraguto,D.,
             Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J.,
             Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,W.,
             Gunaratne,P., Hale,S., Hamilton,K., Han,J., Harris,C., Harris,K.,
             Hart,M., Havlak,P., Hawes,A., Hernandez,J., Hernandez,O.,
             Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F.,
             Howard,S., Huber,J., Hulyk,S., Hume,J., Ioshikhes,I., Jackson,L.E.,
             Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
             Karlsson,E., Kelly,S., Khan,U., King,L., Korval,C., Kovar,C.,
             Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lee,E., Lewis,L.C.,
             Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W.,
             Louiseged,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
             Ma,J., Maheshwari,M., Mapua,P., Marondel,I., Martin,R.,
             Martindale,A., Martinez,E., Massey,E., Mawhiney,E., McLeod,M.P.,
             Meador,M., Mei,G., Mersche,S., Metzker,M., Miller,A., Miner,G.,
             Miner,Z., Mitchell,T., Mohabbat,K., Montgomery,K.T., Morgan,M.,
             Morris,S., Moser,M., Neal,D., Nelson,D., Newton,J., Nelson,N.,
             Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
             Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
```

Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shim,C.,
Shooshtari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A.,
Stanley,H., Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A.,
Tamerisa,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telfrod,B.,
Thomas,N., Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D.,
Vinson,R., Wall,R., Wang,S., Ward-Moore,S., Warren,R.,
Washington,C., Watington,S., Williams,G., Williamson,A.,
Wleczyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zhou,J.,
Zorrilla,S., Kucherlapati,R. and Gibbs,R.

TITLE   Direct Submission
JOURNAL   Unpublished
REFERENCE   2 (bases 1 to 180971)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (27-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3 (bases 1 to 180971)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (01-JAN-2001) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   4 (bases 1 to 180971)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (15-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   5 (bases 1 to 180971)
AUTHORS   Worley,K.C.
TITLE   Direct Submission
JOURNAL   Submitted (27-APR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT   On Jan 1, 2001 this sequence version replaced gi:11968205.
            INFORMATION: http://www.hgsc.bcm.tmc.edu/ or email
            gc-help@bcm.tmc.edu

            CLONE LENGTH: This sequence does not necessarily represent the
            entire insert of this clone. Overlapping regions of clones are only
            sequenced and submitted once, so the sequence for the remainder of
            the insert may be found in the record for the adjacent clones.
            Overlapping clones are noted at the beginning and end of the
            Features listing.

            ANNOTATION OF FEATURES:
            STSs are identified using ePCR (Genome Res. 7:541-550) searches
            of a local database that includes entries from dbSTS, GDB, and
            local mapping efforts.
            Repeats are identified using RepeatMasker (A. Smit and P. Green,
            unpublished.) for Human and Mouse sequences.
            Genes and Region of sequence similarity are identified by BLAST
            (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the
            EST and cDNA sequences. Genes demonstrate at least two exons
            flanked by consensus splice sites that maintained sequence
            continuity across the splice junctions. Sequences that are not
            identical matches are annotated as similar.

            SEQUENCING READ COVERAGE: Sequencing is completed to a minimum
            standard of double strand coverage with a minimum of 2 clones and 2
            reads with no ambiguities or 2 chemistries with a minimum of 2
            clones and 3 reads with no ambiguities. If the sequence quality for
            a region does not meet this standard, it will be indicated in the
            annotation as Low Coverage.

            QUALITY OF INDIVIDUAL BASES:This sequence meets stringent quality
            standards - estimated error rate less than 1 per 10,000 bases.
            Reports of lowest quality individual bases and measures of base
            quality are listed below. Description of the metrics can be found
            at URL:

http://gc.bcm.tmc.edu:8088/quality.info/genbank.annotation.html.

QUALSTAT-REPORT---------

-------- Summary Statistics --------

Contig length: 180971
Phrap values in estimate: 180205
Average error rate (BCM-Phrap estimate):   8.9035e-06
Fraction of Phrap values less than 40 :   0.0075858
Number of consensus changing edits: 50
Number of N's in consensus :   0

-------- Consensus changing edits --------

Position	Original+Context	Edited+Context
445	aaagtgggg(n)aagctgccag	aaagtgggg(g)aagctgccag
5501	tgtccaaaa(n)aaataaataa	tgtccaaaa(t)aataaaataa
5837	tgtaagatat(n)catatatata	tgtaagatat(t)catatata
5855	atatatat(n)tttttttgtt	atatatat(a)tttttttgtt
5881	acagagttta(n)ncttgttgcc	acagagttta(c)tcttgttgcc
5882	cagagttta(n)cttgtgtgcc	cagagttac(t)ctgttgtcc
5953	gggtcaagc(n)atcttctgc	gggttcaagc(a)tcttctgc
5959	aagcaattct(n)tgcctcagc	aagcaatcct(t)ctgcctcagc
5974	ctcagcctcc(n)gagtagctgg	ctcagctcc(n)gagtagctgg
6033	atttttagta(n)agatggggtt	attttagta(g)agatgggtt
6084	cctgacctca(n)gtgntctgcc	cctgacctca(g)tgntctgcc
6088	acctccangtg(n)tctgccnnnn	acctccangtg(t)gntctgcc
6095	gtgntctgcc(n)nnnnnnctc	gtgatctgcc(g)cctggcctc
6096	tgnctgccn(n)nnnnncctc	tgatctgcct(g)cctggcctc
6097	gntctgccnn(n)nnnnncctcc	gatctgcctg(c)ctggcctcc
6098	ntctgcnnn(n)nnnnccctcc	atctgccgc(c)ctggcctcc
6099	tctgccnnn(n)nncctccca	tctgccgcc(t)tggcctcca
6100	ctgccnnnn(n)nncctccaa	ctgcccctt(g)gcctcccaa
6101	tgccnnnnn(n)ncctcccaa	tgcctgcctt(g)gcctcccaag
6102	gcnnnnnnn(n)ctcccaaag	gctgcctg(g)cccccaaag
6735	gttgcagtga(n)ccaagatcgc	gttgcagtga(g)ccaagatcgc
29211	tgcaaaggc(n)tgatgcctgc	tgcaaaggc(c)tgatgcctgc
31019	acgggtttt(n)ccatattggc	acgggttt(c)ccatattggc
31033	tattggccag(n)aaggtcctgn	tattggccag(g)aaggtcctgn
31043	naagtctcg(n)aagtgcattt	naagtctcg(g)aagtgcattt
31083	agtccagaag(n)cccattcct	agtccagaag(c)cccattcct
31124	ctagccctt(n)tatcccaaac	ctagccctt(c)tatcccaaac
31239	taaaagagat(n)cagtctgtcc	taaaagagat(a)cagtctgtcc
32343	gtcagcagtg(n)catagacctt	gtcagcagtg(c)catagacctt
43188	agaaaaaaa(n)ccactctctc	agaaaaaaa(a)ccactctctc
54549	ataggaata(n)aatcaatctg	ataggaata(a)aatcaatctg
68263	gtcaaaagtg(n)ttctctgngn	gtcaaaagtg(t)ttctctgca
68264	tcaaaagtg(n)tctctctgnnc	tcaaaagtg(t)tctctgnnc
68269	agtgnnttct(n)tgnncacat	agtgnnttct(t)tgnncacat
68272	gntcctntg(n)nccacatcac	gntcctntg(c)nccacatcac
68273	nntcctntgn(n)nccacatcac	nntcctntgn(n)nccacatcac
84783	aggttttga(n)ttgnaattgt	aggttttga(a)ttgnaattgt
84787	tttganttg(a)tatgttggt	tttganttg(g)tatgttggt
107633	aactaaatta(n)ncttatgcaat	aactaaatta(a)ttatgcaat
107634	actaattan(n)ttatgcaatc	actaattan(a)ttatgcaatc
119957	tgttggccag(a)ctggtcttga	tgttggccag(g)ctggttga
119521	agctcctcat(n)tgatggaga	agctcctcat(c)tgatggaga
137863	ttaattcctg(n)tccaggtcca	ttaattcctg(a)tccaggtcca
156893	ctaataaaat(n)ttacctgaa	ctaataaaat(c)ttacctgaa
163882	taaattcctg(a)cctgttgaa	taaattcctg(c)cctgttgaa
163892	acctgttga(n)aatctacttc	acctgttga(a)aatctacttc
172348	taaataacag(n)cggcctttaa	taaataacag(c)ggcctttaa
172474	gcagccaaat(n)acaaatgggc	gcagccaaat(a)acaaatgggc
172542	tgtaaaaaac(n)cacctatata	tgtaaaaaac(c)cacctatata
176104	caaagtccag(n)gtatcattct	caaagtccag(t)gtatcattct

-------- Distribution of Quality < 40 Bases --------

```
1000|
 900|
 800|            *   *
```

```
        #   700 |                                                           *
      bases   600 |                                                           *
            500 |                                                           *
            400 |                                                           *
            300 |                                                       *   *
            200 |                                               *   *   *
            100 |                                       *   *   *   *   *   *
              0 |   *   *   *   *   *   *   *   *   *   *   *   *   *   *   *
                -----------------------------------------------------------------
                  5    10   15   20   25   30   35   40
                              Phrap Value Range

Query Match             17.5%;    Score 35.2;  DB 9;   Length 180971;
Best Local Similarity   52.8%;    Pred. No. 4.1;
Matches     76;  Conservative    0;  Mismatches    68;  Indels    0;  Gaps    0;

Qy       21 CCTCTTCCTGCCCACCAGACCAGCAACTTCCAAAATTCTGCCCGTGTTAGCCTTGTT  80
            ||||| |||||||| |||||||| || ||||| || | ||||| |||||||||||||
Db    22236 CCTATCCCTTGCTTACCTGGGATGACTGGTCACTGGTCAGCACGAGGGTGGCAGGGGT  22295

Qy       81 AAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAGGGATCTGATTATTGAGGTGTG  140
            || || |||| |||| ||||||||||| || |||||||||||||||||||| ||||||
Db    22296 GGGGCGGCGCCCTCCGATCTGAGGGAAGGTGGGAAATGGGAAGAGAGGTCAAGTACACGTG  22355

Qy      141 GAAGGGATAAATAAATCAGTCCACA  164
            |||||||| |||||||||||||||
Db    22356 GAAGCATCAAAGCCCCCACCCAAAA  22379

RESULT 15
BX649492
LOCUS        BX649492        189932 bp    DNA       linear   HTG 06-NOV-2003
DEFINITION   Danio rerio clone DKEY-13P3, WORKING DRAFT SEQUENCE, 2 unordered
             pieces.
ACCESSION    BX649492
VERSION      BX649492.4  GI:38091215
KEYWORDS     HTG; HTGS_PHASE1; HTGS_ACTIVEFIN; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE       Danio rerio (zebrafish)
ORGANISM     Danio rerio
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
             Cypriniformes; Cyprinidae; Danio.
REFERENCE    1 (bases 1 to 189932)
AUTHORS      Bates,K.
TITLE        Direct Submission
JOURNAL      Submitted (05-NOV-2003) Wellcome Trust Sanger Institute, Hinxton,
             Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
             zfish-help@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk
             On Oct 30, 2003 this sequence version replaced gi:37988067.
             ------------------- Genome Center
             Center: Wellcome Trust Sanger Institute
             Center code: SC
             Web site: http://www.sanger.ac.uk
             Contact: zfish-help@sanger.ac.uk
             ------------------- Project Information
             Center project name: zK13P3
             ------------------- Summary Statistics
             Assembly program: XGAP4; version 4.5
             Chemistry: Dye-terminator; 100% of reads
             Consensus quality: 189785 bases at least Q40
             Consensus quality: 189794 bases at least Q30
             Consensus quality: 189816 bases at least Q20
             Insert size: 189832; sum-of-contigs
             Insert size: 193531; 3.6% error; agarose-fp
             Quality coverage: 19.57x in Q20 bases; sum-of-contigs Quality
             coverage: 20.58x in Q20 bases; agarose-fp
             -------------------
             * NOTE: This is a 'working draft' sequence. It currently
             * consists of 2 contigs. The true order of the pieces
             * is not known and their order in this sequence record is
             * arbitrary. Gaps between the contigs are represented as
             * runs of N, but the exact sizes of the gaps are unknown.
             * This record will be updated with the finished sequence
             * as soon as it is available and the accession number will
             * be preserved.
             *      1  112060: contig of 112060 bp in length
             * 112060  112160: gap of 100 bp
             * 112161  189932: contig of 77772 bp in length.
FEATURES             Location/Qualifiers
     source          1..189932
                     /organism="Danio rerio"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7955"
                     /clone="DKEY-13P3"
                     /clone_lib="DanioKey"
     misc_feature    1..112060
                     /note="assembly_fragment:03464.0"
     misc_feature    112161..189932
                     /note="assembly_fragment:05810"
ORIGIN

Query Match             17.5%;    Score 35.2;  DB 2;   Length 189932;
Best Local Similarity   53.7%;    Pred. No. 4.1;
Matches     73;  Conservative    0;  Mismatches   63;  Indels    0;  Gaps    0;

Qy       56 AATTCTGCCCGTGTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAA  115
            ||||||| ||| |||||| || ||||| ||||  |||| |||||||||||||||||||
Db    64914 AATTATGCTGCTGATTAGCCTAATTAAAGCTTAGCTTATTCCACTTAAGGATTAGGGGCTA  64973

Qy      116 AAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTCCACAAATAAACAAAC  175
            ||||| ||||||||| |||||||||| |||| |||||||| |||| ||| |||||| ||
Db    64974 AAGTAAGCCCGATTATAAATGTGAGTGTGTTCAGAAATCATAACTGTACAATTCCCTGGTC  65033

Qy      176 TGTCCGGGATTCCTAG  191
            ||||||||||||| ||
Db    65034 TGCCCATGAGACAGAG  65049

Search completed: March 25, 2004, 12:50:05
Job time : 1375.52 secs
```

```
                 GenCore version 5.1.6
         Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:     March 25, 2004, 07:40:23 ; Search time 208.591 Seconds
                                        (without alignments)
                                        4093.601 Million cell updates/sec

Title:      US-09-963-285-1_COPY_223_423
Perfect score: 201
Sequence:   1 acaaatgttctccctgaagc.........ggatcctagaggggaaggag 201

Scoring table: IDENTITY.NUC
            Gapop 10.0 , Gapext 1.0

Searched:   3373863 seqs, 2124099041 residues

Total number of hits satisfying chosen parameters:    6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  N_Geneseq_29Jan04:*
            1: geneseqn1980s:*
            2: geneseqn1990s:*
            3: geneseqn2000s:*
            4: geneseqn2001as:*
            5: geneseqn2001bs:*
            6: geneseqn2002s:*
            7: geneseqn2003as:*
            8: geneseqn2003bs:*
            9: geneseqn2003cs:*
            10: geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	201	100.0	6458	6	ABK48984	Abk48984 Genomic D
2	66	32.8	6021	6	ABK48986	Abk48986 Genomic D
3	32.4	16.1	6962	7	ACC90589	Acc90589 Human CGD
4	31.6	15.7	5983	4	AAK84657	Aak84657 Human imm
5	31.4	15.6	296	7	ABX81870	Abx81870 Corn ear-
6	31.4	15.6	22700	4	AAK70746	Aak70746 Human imm
7	31.2	15.5	213456	7	AAL54072	Aal54072 Genomic D
8	31	15.4	1830	6	ABZ11734	Abz11734 Human pol
9	31	15.4	110000	7	AAD53223_3	Continuation (4 of
c 10	30.8	15.3	360	4	ABX41793	Abx41793 Bovine ES
11	30.4	15.1	491	4	ABS30387	Abs30387 Human liv
c 12	30.4	15.1	3574	2	AAX56241	Aax56241 Mouse JMY
c 13	30.2	15.0	1466	3	AAA72296	Aaa72296 Drosophil
14	30.2	15.0	2075	3	AAA94812	Aaa94812 Drosophil
c 15	30.2	15.0	3548	4	ABL22180	Abl22180 Drosophil
c 16	30.2	15.0	9303	7	ABZ67485	Abz67485 Human sec
c 17	30.2	15.0	9303	7	ABZ73903	Abz73903 Secreted
18	30	14.9	4466	4	AAK69983	Aak69983 Human imm
19	29.8	14.8	274	7	ABX86185	Abx86185 Corn ear-
20	29.6	14.7	816	6	ABN68994	Abn68994 Streptoco
c 21	29.6	14.7	110000	6	ABN71527_12	Continuation (13 o
c 22	29.6	14.7	142229	6	ADD50651	Add50651 BAC seque
c 23	29.4	14.6	199	3	AAC21229	Aac21229 Human sec
c 24	29.4	14.6	336	5	ABV03952	Abv03952 Human pro
c 25	29.4	14.6	2103	9	ADB63634	Adb63634 Human cDN
c 26	29.4	14.6	2960	4	ABL05844	Abl05844 Drosophil
27	29.4	14.6	32885	4	ABL19560	Abl19560 Drosophil
c 28	29.4	14.6	40681	6	ABA92787_6	Continuation (7 of
c 29	29.4	14.6	63115	9	ADC85419	Adc85419 Mouse Tle
c 30	29.2	14.5	2606	4	AAH17011	Aah17011 Human cDN
31	29.2	14.5	6162	6	ABK31331	Abk31331 Signal tr
c 32	29.2	14.5	6943	9	ADC86776	Adc86776 Human GPC
33	29.2	14.5	15004	4	AAH27885	Aah27885 Nucleotid
c 34	29.2	14.5	110000	7	AAV21209_07	Continuation (8 of
c 35	29	14.4	900	3	AAC46407	Aac46407 Arabidops
c 36	29	14.4	900	3	AAC37057	Aac37057 Arabidops
c 37	29	14.4	2000	7	ADA71938	Ada71938 Rice gene
38	29	14.4	7058	6	ABK40051	Abk40051 Human che
39	29	14.4	9469	5	AAK79514	Aak79514 Human imm
c 40	29	14.4	177563	8	ACD28257	Acd28257 Mouse sol
c 41	29	14.4	222880	9	ADC87622	Adc87622 Human GPC
c 42	28.8	14.3	417	7	AAC07240	Aac07240 Human sec
c 43	28.8	14.3	1287	7	ACC61427	Acc61427 Gene sequ
c 44	28.8	14.3	1531	4	AAH34920	Aah34920 Human col
c 45	28.8	14.3	1948	6	ABK34716	Abk34716 Human cDN

ALIGNMENTS

```
RESULT 1
ABK48984
ID   ABK48984 standard; DNA; 6458 BP.
XX
AC   ABK48984;
XX
DT   02-JUL-2002  (first entry)
XX
DE   Genomic DNA encoding human transcription factor FOXC2.
XX
KW   Transcription factor; FOXC2; antidiabetic; anorectic; antilipaemic;
KW   cardiovascular; FOXC2 polypeptide expression modulator; FOXC2 promoter;
KW   diabetes mellitus; obesity; hypercholesterolaemia; dyslipidaemia;
KW   cardiovascular disease; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FH
FT   exon            1..215
FT                   /*tag= b
FT                   /note= "First exon according to the alternative
FT                   transcript"
FT   5'UTR           1..186
FT                   /*tag= a
FT   misc_feature    187..215
FT                   /*tag= c
FT                   /note= "Region coding for 5'part of alternative protein"
FT   misc_signal     215..216
FT                   /*tag= d
FT                   /note= "Alternative first exon splice site"
FT   misc_difference 216..475
FT                   /*tag= e
FT                   /note= "Fragment of the FOXC2 enhancer. Specifically
FT                   claimed in claim 16"
FT   misc_difference 223..231
FT                   /*tag= f
FT                   /note= "Fragment of the FOXC2 enhancer. Specifically
FT                   claimed in claim 12"
FT   misc_difference 359..375
FT                   /*tag= g
FT                   /note= "Fragment of the FOXC2 enhancer. Specifically
FT                   claimed in claim 13"
FT   misc_difference 378..402
FT                   /*tag= h
FT                   /note= "Fragment of the FOXC2 enhancer. Specifically
FT                   claimed in claim 14"
```

```
Best Local Similarity   100.0%;  Pred. No. 3.8e-58;
Matches 201;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 ACAAATGTCTCCCTGAAGCCCTCTTCCCTGCCCAACCAGACCAGCAACTTCCAAAATTC 60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   223 ACAAATGTCTCCCTGAAGCCCTCTTCCCTGCCCAACCAGACCAGCAACTTCCAAAATTC 282

Qy    61 TGCCCGTGTTGAGCCTTGTTAAAGGGGTGTTCTCTCACTCCTTCAGGGAAAGTGGGAAAAGGG 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   283 TGCCCGTGTTGAGCCTTGTTAAAGGGGTGTTCTCTCACTCCTTCAGGGAAAGTGGGAAAAGGG 342

Qy   121 GATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTCCACAAATAAACAAACTGTCC 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   343 GATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTCCACAAATAAACAAACTGTCC 402

Qy   181 GGGATTCCTAGAGGGAAGGAG 201
         |||||||||||||||||||||
Db   403 GGGATTCCTAGAGGGAAGGAG 423


RESULT 2
ABK48986
ID   ABK48986 standard; DNA; 6021 BP.
XX
AC   ABK48986;
XX
DT   02-JUL-2002  (first entry)
XX
DE   Genomic DNA encoding mouse mesenchyme forkhead-1 (MHF-1)/FOXC2.
XX
KW   Transcription factor; FOXC2; antidiabetic; anorectic; antilipaemic;
KW   cardiovascular; FOXC2 polypeptide expression modulator; FOXC2 promoter;
KW   diabetes mellitus; obesity; hypercholesterolaemia; dyslipidaemia;
KW   cardiovascular disease; mouse mesenchyme forkhead 1; MHF-1; gene; ds.
XX
OS   Mus musculus.
XX
FH   Key             Location/Qualifiers
FT   CDS             2070..3554
FT                   /*tag= a
FT                   /product= "MHF-1 or FOXC2"
FT                   /note= "Mouse mesenchyme forkhead 1"
XX
PN   WO20027008-A1.
XX
PD   04-APR-2002.
XX
PF   26-SEP-2001; 2001WO-SE002098.
XX
PR   26-SEP-2000; 2000SE-00003435.
PR   10-OCT-2000; 2000US-0238897P.
PR   09-NOV-2000; 2000SE-00004102.
XX
PA   (BIOV-) BIOVITRUM AB.
XX
PI   Enerbaeck S, Krook K, Rondahl L, Wasserman WW;
XX
DR   WPI; 2002-352129/38.
DR   P-PSDB; AAU79818.
XX
PT   An isolated FOXC2 promoter region that modulates the expression of a
PT   FOXC2 polypeptide is useful for treating obesity and type II diabetes
PT   mellitus.
XX
PS   Claim 22; Page 45-46; 62pp; English.
XX
CC   The invention describes an isolated human FOXC2 promoter region. The
CC   mammalian FOXC2 promoter region for the modulation of FOXC2 polypeptide
CC   expression has therapeutic value in treating type II diabetes mellitus,
CC   obesity, hypercholesterolaemia, other cardiovascular diseases or
CC   dyslipidaemias. This sequence encodes the mouse mesenchyme forkhead 1
CC   (MHF-1) protein (also called FOXC2 transcription factor), described in
CC   the method of the invention
```

```
FT   misc_difference 403..423
FT                   /*tag= i
FT                   /note= "Fragment of the FOXC2 enhancer. Specifically
FT                   claimed in claim 15"
FT   misc_difference 1250..2235
FT                   /*tag= k
FT                   /note= "Fragment of the FOXC2 promoter. Specifically
FT                   claimed in claim 3"
FT   misc_difference 1250..1749
FT                   /*tag= j
FT                   /note= "Fragment of the FOXC2 promoter. Specifically
FT                   claimed in claim 2"
FT   misc_difference 1692..1703
FT                   /*tag= l
FT                   /note= "Fragment of the FOXC2 promoter. Specifically
FT                   claimed in claim 1"
FT   exon            1746..4629
FT                   /*tag= n
FT                   /note= "First exon according to the published form of the
FT                   transcript"
FT   5'UTR           1746..2234
FT                   /*tag= m
FT   CDS             2235..3740
FT                   /*tag= o
FT                   /product= "FOXC2"
FT                   /note= "Transcription factor"
FT   misc_binding    2448..2735
FT                   /*tag= p
FT                   /note= "Region coding for DNA-binding domain"
FT   exon            2516..4629
FT                   /*tag= r
FT                   /note= "Second exon according to the alternative
FT                   transcript"
FT   misc_feature    2516..3740
FT                   /*tag= q
FT                   /note= "Portion of polypeptide used in alternative
FT                   transcript"
FT   3'UTR           3741..4629
FT                   /*tag= s
XX
PN   WO20027008-A1.
XX
PD   04-APR-2002.
XX
PF   26-SEP-2001; 2001WO-SE002098.
XX
PR   26-SEP-2000; 2000SE-00003435.
PR   10-OCT-2000; 2000US-0238897P.
PR   09-NOV-2000; 2000SE-00004102.
XX
PA   (BIOV-) BIOVITRUM AB.
XX
PI   Enerbaeck S, Krook K, Rondahl L, Wasserman WW;
XX
DR   WPI; 2002-352129/38.
DR   P-PSDB; AAU79816.
XX
PT   An isolated FOXC2 promoter region that modulates the expression of a
PT   FOXC2 polypeptide is useful for treating obesity and type II diabetes
PT   mellitus.
XX
PS   Claim 3; Page 33-38; 62pp; English.
XX
CC   The invention describes an isolated human FOXC2 promoter region. The
CC   mammalian FOXC2 promoter region for the modulation of FOXC2 polypeptide
CC   expression has therapeutic value in treating type II diabetes mellitus,
CC   obesity, hypercholesterolaemia, other cardiovascular diseases or
CC   dyslipidaemias. This sequence encodes the human transcription factor
CC   FOXC2, the promoter of which is described in the invention
XX
SQ   Sequence 6458 BP; 1443 A; 1826 C; 1715 G; 1474 T; 0 U; 0 Other;

Query Match    100.0%; Score 201; DB 6; Length 6458;
```

```
XX
SQ   Sequence 6021 BP; 1350 A; 1675 C; 1642 G; 1354 T; 0 U; 0 Other;

Query Match     32.8%;  Score 66; DB 6;  Length 6021;
Best Local Similarity   73.1%;  Pred. No. 1e-11;
Matches 155;  Conservative   0;  Mismatches  45;  Indels  12;  Gaps    5;

Qy       1 ACAAATGTTCTCCTGAAGCCCTTCCTGCCAACCAGACCAGCAACTTCCAAAATTC  60
           ||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db     195 ACAAATGTTTATCTGTCGCCCTTCTTCCCTACCCAACCAACAACTTCCAGAAGGT 254

Qy      61 TGCCCGTG--TTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAA 117
           |||| ||    |||||  ||||| |||||||| |||||||||  |||||  ||||| |
Db     255 TCTGCGGAGGCATAGAGCCATTCCGTAGGGACATCTCGGTGCTTCTGAGGAGCGGACCGA 314

Qy     118 G--GGGATCTGATTA---TTGAGGTGTGGGAAGGAATAAATAATCAGTCCACAAATAAAC 171
           |  ||||||||||||      |||| ||||||||||||||  |||||||||||||||||
Db     315 GCAGGGATCCGATGACGACTGGAGATGTTGAAGGAATAAAT-ACCAGTCCACAAATAAAC 373

Qy     172 AAACTGT--CCGGGATCCTAGAGGGAAGGAG 201
           ||||||    |||||||||||||||||||||
Db     374 AAACTGTCCCCGGGATTCCTAGAGGGAAGGAG 405

RESULT 3
ACC90589
ID   ACC90589 standard; DNA; 6962 BP.
XX
AC   ACC90589;
XX
DT   12-AUG-2003 (first entry)
XX
DE   Human CGDD-12 encoding DNA.
XX
KW   Human; cytostatic; antiarteriosclerotic; anticonvulsant; nootropic;
KW   neuroprotective; cerebroprotective; anti-HIV; antiallergic;
KW   antiinflammatory; gynaecological; cancer; atherosclerosis; epilepsy;
KW   Huntington's disease; stroke; AIDS; allergy; placenta; reproductive;
KW   CGDD; cell growth; cell differentiation; cell death; gene; ds.
XX
OS   Homo sapiens.
XX
PN   WO2003027263-A2.
XX
PD   03-APR-2003.
XX
PF   26-SEP-2002; 2002WO-US031095.
XX
PR   28-SEP-2001; 2001US-0326389P.
PR   05-OCT-2001; 2001US-0327380P.
PR   05-OCT-2001; 2001US-0328186P.
PR   05-OCT-2001; 2001US-0329690P.
PR   12-OCT-2001; 2001US-0345384P.
PR   26-OCT-2001; 2001US-0348165P.
PR   02-NOV-2001; 2001US-0350219P.
PR   09-NOV-2001; 2001US-0344518P.
PR   09-NOV-2001; 2001US-0345143P.
PR   16-NOV-2001; 2001US-0332375P.
PR   03-DEC-2001; 2001US-0336908P.
PR   07-DEC-2001; 2001US-0340747P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Azimzai Y,    Baughn MR,   Becha SD,   Borowsky ML,   Chawla NK;
PI   Elliott VS,   Emerling BM, Gandhi AR,  Gietzen KJ,    Gorvad AE;
PI   Griffin JA,   Hafalia AJA, Ison CH,    Kable AE,      Kalafus DP;
PI   Lehr-Mason PM, Lu DAM,     Marquis JP, Nguyen DB,     Ramkumar J;
PI   Richardson TW, Sapperstein SK, Swarnakar A,  Tang YT,  Tran UK;
PI   Warren BA,    Xu Y,       Yao MG,     Yue H,         Yue H;
XX
DR   WPI; 2003-421159/39.
DR   P-PSDB; ABR69612.
XX
```

```
PT   New human proteins associated with cell growth, differentiation, and
PT   death (CGDD), useful for diagnosing, treating and preventing diseases or
PT   conditions associated with the aberrant CGDD expression e.g. cancer,
PT   AIDS, or epilepsy.
XX
PS   Claim 5; Page 317-318; 350pp; English.
XX
CC   The invention relates to an isolated polypeptide associated with cell
CC   growth, differentiation and death (CGDD). Also disclosed are the
CC   polynucleotides encoding the polypeptides. The polypeptides and
CC   polynucleotides are useful in diagnosing, treating and preventing
CC   diseases or conditions associated with the decreased expression or over
CC   expression of CGDD. Such diseases include cell proliferative (e.g.
CC   cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's
CC   disease, stroke), immune/inflammatory (e.g. AIDS, allergies) and
CC   reproductive disorders, or disorders of the placenta. They are also
CC   useful in assessing the effects of exogenous compounds on the expression
CC   of nucleic acid and amino acid sequences of CGDD. The CGDD or its
CC   fragments are useful in screening compounds for effectiveness as an
CC   agonist or antagonist of the polypeptides, or in altering the expression
CC   of the target polynucleotide and compounds that specifically bind to or
CC   modulate the activity of the polypeptide. Microarrays consisting
CC   polynucleotides of the invention are useful in monitoring or measuring
CC   protein-protein interactions, drug-target interactions, and gene
CC   expression profiles. Sequences given in records ACC90578-ACC90634
CC   represent polynucleotides encoding CGDD proteins of the invention
XX
SQ   Sequence 6962 BP; 2330 A; 1246 C; 1573 G; 1813 T; 0 U; 0 Other;

Query Match     16.1%;  Score 32.4; DB 7;  Length 6962;
Best Local Similarity   56.6%;  Pred. No. 3.8;
Matches  60;  Conservative   0;  Mismatches  46;  Indels   0;  Gaps    0;

Qy      73 GCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAGTGGGAAAAGGGGATCTGATTATT 132
           |||||||  |||||||||  ||||||||||||||||||||| |||||||||  ||| ||
Db    6835 GCTTTTTTTAAGGGGTTCAGATCCTGGAAAGAAAATGCACAATAGTGACTATGAGAATG 6894

Qy     133 GAGGTGTGGAAGGAATAAATAATCAGTCCACAAATAAACAAACTGT 178
           |||||||||||||||||||||||||| ||||||||||||||
Db    6895 TAGAAGTGGCATCTGAAAACAATTGGTACACAAGTAAACTGGCAGT 6940

RESULT 4
AAK84657
ID   AAK84657 standard; DNA; 5983 BP.
XX
AC   AAK84657;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:39469.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO20015718Z-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US001354.
XX
PR   31-JAN-2000; 2000US-0179065P.
PR   04-FEB-2000; 2000US-0180628P.
PR   24-FEB-2000; 2000US-0184664P.
PR   02-MAR-2000; 2000US-0186350P.
PR   16-MAR-2000; 2000US-0189974P.
PR   17-MAR-2000; 2000US-0190076P.
PR   18-APR-2000; 2000US-0198123P.
PR   19-MAY-2000; 2000US-0205515P.
PR   07-JUN-2000; 2000US-0209467P.
PR   28-JUN-2000; 2000US-0214886P.
```

PR	30-JUN-2000;	2000US-0215135P.
PR	07-JUL-2000;	2000US-0216647P.
PR	07-JUL-2000;	2000US-0216880P.
PR	11-JUL-2000;	2000US-0217487P.
PR	11-JUL-2000;	2000US-0217496P.
PR	14-JUL-2000;	2000US-0218290P.
PR	26-JUL-2000;	2000US-0220963P.
PR	26-JUL-2000;	2000US-0220964P.
PR	14-AUG-2000;	2000US-0224518P.
PR	14-AUG-2000;	2000US-0224519P.
PR	14-AUG-2000;	2000US-0225213P.
PR	14-AUG-2000;	2000US-0225214P.
PR	14-AUG-2000;	2000US-0225266P.
PR	14-AUG-2000;	2000US-0225267P.
PR	14-AUG-2000;	2000US-0225268P.
PR	14-AUG-2000;	2000US-0225270P.
PR	14-AUG-2000;	2000US-0225447P.
PR	14-AUG-2000;	2000US-0225757P.
PR	14-AUG-2000;	2000US-0225758P.
PR	14-AUG-2000;	2000US-0225759P.
PR	18-AUG-2000;	2000US-0226279P.
PR	22-AUG-2000;	2000US-0226681P.
PR	22-AUG-2000;	2000US-0226868P.
PR	22-AUG-2000;	2000US-0227182P.
PR	23-AUG-2000;	2000US-0227009P.
PR	30-AUG-2000;	2000US-0228924P.
PR	01-SEP-2000;	2000US-0229287P.
PR	01-SEP-2000;	2000US-0229344P.
PR	01-SEP-2000;	2000US-0229345P.
PR	05-SEP-2000;	2000US-0229509P.
PR	05-SEP-2000;	2000US-0229513P.
PR	06-SEP-2000;	2000US-0230437P.
PR	08-SEP-2000;	2000US-0230438P.
PR	08-SEP-2000;	2000US-0231242P.
PR	08-SEP-2000;	2000US-0231243P.
PR	08-SEP-2000;	2000US-0231244P.
PR	08-SEP-2000;	2000US-0231413P.
PR	08-SEP-2000;	2000US-0231414P.
PR	08-SEP-2000;	2000US-0232080P.
PR	08-SEP-2000;	2000US-0232081P.
PR	12-SEP-2000;	2000US-0231968P.
PR	14-SEP-2000;	2000US-0232397P.
PR	14-SEP-2000;	2000US-0232398P.
PR	14-SEP-2000;	2000US-0232399P.
PR	14-SEP-2000;	2000US-0232400P.
PR	14-SEP-2000;	2000US-0232401P.
PR	14-SEP-2000;	2000US-0233063P.
PR	14-SEP-2000;	2000US-0233064P.
PR	14-SEP-2000;	2000US-0233065P.
PR	21-SEP-2000;	2000US-0234223P.
PR	21-SEP-2000;	2000US-0234274P.
PR	25-SEP-2000;	2000US-0234937P.
PR	25-SEP-2000;	2000US-0234998P.
PR	26-SEP-2000;	2000US-0235484P.
PR	27-SEP-2000;	2000US-0235834P.
PR	27-SEP-2000;	2000US-0235835P.
PR	29-SEP-2000;	2000US-0236327P.
PR	29-SEP-2000;	2000US-0236367P.
PR	29-SEP-2000;	2000US-0236368P.
PR	29-SEP-2000;	2000US-0236369P.
PR	29-SEP-2000;	2000US-0236370P.
PR	13-OCT-2000;	2000US-0239935P.
PR	13-OCT-2000;	2000US-0239937P.
PR	02-OCT-2000;	2000US-0236802P.
PR	02-OCT-2000;	2000US-0237037P.
PR	02-OCT-2000;	2000US-0237038P.
PR	02-OCT-2000;	2000US-0237039P.
PR	02-OCT-2000;	2000US-0237040P.
PR	20-OCT-2000;	2000US-0240960P.
PR	20-OCT-2000;	2000US-0241221P.
PR	20-OCT-2000;	2000US-0241785P.
PR	20-OCT-2000;	2000US-0241786P.
PR	20-OCT-2000;	2000US-0241787P.
PR	20-OCT-2000;	2000US-0241808P.
PR	20-OCT-2000;	2000US-0241809P.
PR	20-OCT-2000;	2000US-0241826P.
PR	01-NOV-2000;	2000US-0246617P.
PR	08-NOV-2000;	2000US-0246474P.
PR	08-NOV-2000;	2000US-0246475P.
PR	08-NOV-2000;	2000US-0246476P.
PR	08-NOV-2000;	2000US-0246477P.
PR	08-NOV-2000;	2000US-0246478P.
PR	08-NOV-2000;	2000US-0246523P.
PR	08-NOV-2000;	2000US-0246524P.
PR	08-NOV-2000;	2000US-0246525P.
PR	08-NOV-2000;	2000US-0246526P.
PR	08-NOV-2000;	2000US-0246527P.
PR	08-NOV-2000;	2000US-0246528P.
PR	08-NOV-2000;	2000US-0246532P.
PR	08-NOV-2000;	2000US-0246609P.
PR	08-NOV-2000;	2000US-0246610P.
PR	08-NOV-2000;	2000US-0246611P.
PR	08-NOV-2000;	2000US-0246613P.
PR	17-NOV-2000;	2000US-0249207P.
PR	17-NOV-2000;	2000US-0249208P.
PR	17-NOV-2000;	2000US-0249209P.
PR	17-NOV-2000;	2000US-0249210P.
PR	17-NOV-2000;	2000US-0249211P.
PR	17-NOV-2000;	2000US-0249212P.
PR	17-NOV-2000;	2000US-0249213P.
PR	17-NOV-2000;	2000US-0249214P.
PR	17-NOV-2000;	2000US-0249215P.
PR	17-NOV-2000;	2000US-0249216P.
PR	17-NOV-2000;	2000US-0249217P.
PR	17-NOV-2000;	2000US-0249218P.
PR	17-NOV-2000;	2000US-0249244P.
PR	17-NOV-2000;	2000US-0249245P.
PR	17-NOV-2000;	2000US-0249264P.
PR	17-NOV-2000;	2000US-0249265P.
PR	17-NOV-2000;	2000US-0292997P.
PR	17-NOV-2000;	2000US-0292999P.
PR	17-NOV-2000;	2000US-0293000P.
PR	01-DEC-2000;	2000US-0250160P.
PR	01-DEC-2000;	2000US-0250391P.
PR	05-DEC-2000;	2000US-0251030P.
PR	05-DEC-2000;	2000US-0251988P.
PR	05-DEC-2000;	2000US-0257719P.
PR	06-DEC-2000;	2000US-0251479P.
PR	08-DEC-2000;	2000US-0251856P.
PR	08-DEC-2000;	2000US-0251868P.
PR	08-DEC-2000;	2000US-0251869P.
PR	08-DEC-2000;	2000US-0251989P.
PR	08-DEC-2000;	2000US-0251990P.
PR	11-DEC-2000;	2000US-0254097P.
PR	05-JAN-2001;	2001US-0259678P.
XX		
PA	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
XX		
DR	WPI; 2001-48426/52.	
XX		
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating cancers and metastasis.	
XX		
PS	Disclosure; SEQ ID NO 39469; 3071pp + Sequence Listing; English.	
XX		
CC	AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)	
CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic	
CC	activity, and can be used in gene therapy and vaccine production. (I)	
CC	proteins and polynucleotides may be used in the prevention, diagnosis and	
CC	treatment of diseases associated with inappropriate (I) expression. For	
CC	example, they may be used to treat disorders associated with decreased	
CC	expression by rectifying mutations or deletions in a patient's genome	
CC	that affect the activity of (I) by expressing inactive proteins or to	

```
CC    supplement the patients own production of (I). Additionally, (I)
CC    polynucleotides may be used to produce the secreted (I), by inserting the
CC    nucleic acids into a host cell and culturing the cell to express the
CC    protein. (I) proteins and polynucleotides may be used to prevent,
CC    diagnose and treat immune/haematopoietic-related diseases, especially
CC    cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC    to AAK87694 represent human immune/haematopoietic antigen genomic
CC    sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC    represent sequences used in the exemplification of the present invention
XX
SQ    Sequence 5983 BP; 1422 A; 1843 C; 1514 G; 1204 T; 0 U; 0 Other;

Query Match     15.7%;  Score 31.6;  DB 4;  Length 5983;
Best Local Similarity   56.9%;  Pred. No. 6.7;
Matches   58;  Conservative   0;  Mismatches  44;  Indels   0;  Gaps   0;

Qy         2 CAAATGTCTCCCTGAAGCCCTTCCCTGCCAACCAGACCAACTTCCAAAATTCT 61
             ||| ||||||| |||||||  ||||||| |||||||| |||||| |||||||||
Db       772 CAAAAATCCCTCAGTTCTACTCCCTCCATTCCCATGCAGCCCCACCTCCTCCCCACCACA 831

Qy        62 GCCGTGTTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAG 103
             ||||| |||||||||||||||  |||||| ||||| ||||||
Db       832 GCCCCTGGTTCTTCTTGGCCTGAATGAGCGTCTCCCACCTCCCTG 873
```

RESULT 5
ABX81870
```
ID   ABX81870 standard; cDNA; 296 BP.
XX
AC   ABX81870;
XX
DT   24-APR-2003 (first entry)
XX
DE   Corn ear-derived polynucleotide (cpd) #330.
XX
KW   Corn ear-derived polynucleotide; cdp; cDNA library; SATMON022; SATMON023;
KW   structural gene; functional gene; regulatory gene;
KW   corn ear-specific profile; gene transcription; gene expression;
KW   hybrid plant; desirable trait expression; plant breeding program;
KW   inheritance; desired characteristic; growth; development;
KW   disease resistance; environmental adaptability; quality; yield;
KW   multigene trait; plant; gene; ss.
XX
OS   Zea mays.
XX
PN   US6476212-B1.
XX
PD   05-NOV-2002.
XX
PF   14-MAY-1999;    99US-00313294.
XX
PR   26-MAY-1998;    98US-0086722P.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Lalgudi RV,  Ito LY,  Sherman BK;
XX
DR   WPI; 2003-208840/20.
XX
PT   Novel purified corn-ear derived polynucleotide useful as hybridization
PT   probe for detecting polynucleotide in sample, and for identifying,
PT   evaluating, and altering desired characteristics associated with growth,
PT   development.
XX
PS   Example; SEQ ID NO 330; 390pp; English.
XX
CC   The present invention relates to the isolation of corn ear-derived
CC   polynucleotides (cdps) from the corn (Zea mays) cDNA libraries SATMON022
CC   and SATMON023. Some of the cpds uniquely identify structural, functional,
CC   and regulatory genes of corn ear. The polynucleotides sequences are
CC   useful for detecting cpds in a sample, for producing a corn ear-specific
CC   profile of gene transcription, for detecting altered gene expression in
CC   inbred or hybrid plants, and for screening several molecules for specific
```

```
CC    binding to the polynucleotide. The cdps are useful to identify, isolate,
CC    or extend identical or related corn-ear nucleic acid sequences from DNA
CC    libraries, and in nucleic acid amplification or hybridisation techniques
CC    to follow the expression of desirable traits through plant breeding
CC    programs. Preferably, the cdps are used to identify, evaluate, alter, or
CC    follow the inheritance of desired characteristics associated with growth
CC    and development, disease resistance, environmental adaptability, quality,
CC    and yield of corn. The cdps are also useful as molecular markers for
CC    studying inheritance and multigene traits in a plant breeding program.
CC    The cpds are useful for producing purified corn-ear polypeptides by
CC    recombinant techniques. They are also useful in diagnostic assays to
CC    detect or confirm conditions or diseases associated with abnormal levels
CC    of cdp expression. ABX81541-ABX89140 represent corn ear-derived
CC    polynucleotides (cpds) of the invention. Note: The sequence data for this
CC    patent did not form part of the printed specification, but was obtained
CC    in electronic format directly from the USPTO web site at
CC    seqdata.uspto.gov/psipsDIEntry.html
XX
SQ    Sequence 296 BP; 73 A; 59 C; 83 G; 78 T; 0 U; 3 Other;

Query Match     15.6%;  Score 31.4;  DB 7;  Length 296;
Best Local Similarity   55.7%;  Pred. No. 2.1;
Matches   59;  Conservative   0;  Mismatches  47;  Indels   0;  Gaps   0;

Qy        91 CTCACTCCTTCAGGGAAAGTGGGAAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAA 150
             || ||||  |||| |||||  |||||| ||| || ||| |||| |||||||||||||||
Db        76 CTTACATACACCGGTCAGGTCGGACAGGAGAGCAGGTAATACTGGGTGTTGCTGTCATGC 135

Qy       151 ATAATCAGTCCACAAATAAACAAACTGTCCGGGATTCCTAGAGGGA 196
             ||||  ||||| ||||||| ||||| |||| |||| || |||||||
Db       136 TTTATGAGCCCAGATATAAGCACAGTGTCAGNAGACTAGAAGGGA 181
```

RESULT 6
AAK70746
```
ID   AAK70746 standard; DNA; 22700 BP.
XX
AC   AAK70746;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:2558.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001;  2001WO-US001354.
XX
PR   31-JAN-2000;  2000US-0179065P.
PR   04-FEB-2000;  2000US-0180628P.
PR   24-FEB-2000;  2000US-0184664P.
PR   02-MAR-2000;  2000US-0186350P.
PR   16-MAR-2000;  2000US-0189874P.
PR   17-MAR-2000;  2000US-0190076P.
PR   18-APR-2000;  2000US-0198123P.
PR   19-MAY-2000;  2000US-0205515P.
PR   07-JUN-2000;  2000US-0209467P.
PR   28-JUN-2000;  2000US-0214886P.
PR   30-JUN-2000;  2000US-0215135P.
PR   07-JUL-2000;  2000US-0216647P.
PR   07-JUL-2000;  2000US-0216880P.
PR   11-JUL-2000;  2000US-0217487P.
PR   11-JUL-2000;  2000US-0274967P.
PR   14-JUL-2000;  2000US-0218290P.
PR   26-JUL-2000;  2000US-0220963P.
PR   26-JUL-2000;  2000US-0220964P.
PR   14-AUG-2000;  2000US-0224518P.
```

PR   14-AUG-2000;   2000US-0224519P.
PR   14-AUG-2000;   2000US-0225213P.
PR   14-AUG-2000;   2000US-0225214P.
PR   14-AUG-2000;   2000US-0225266P.
PR   14-AUG-2000;   2000US-0225267P.
PR   14-AUG-2000;   2000US-0225268P.
PR   14-AUG-2000;   2000US-0225270P.
PR   14-AUG-2000;   2000US-0225447P.
PR   14-AUG-2000;   2000US-0225757P.
PR   14-AUG-2000;   2000US-0225758P.
PR   14-AUG-2000;   2000US-0225759P.
PR   18-AUG-2000;   2000US-0226279P.
PR   22-AUG-2000;   2000US-0226681P.
PR   22-AUG-2000;   2000US-0226868P.
PR   22-AUG-2000;   2000US-0227182P.
PR   23-AUG-2000;   2000US-0227009P.
PR   30-AUG-2000;   2000US-0228924P.
PR   01-SEP-2000;   2000US-0229287P.
PR   01-SEP-2000;   2000US-0229343P.
PR   01-SEP-2000;   2000US-0229344P.
PR   01-SEP-2000;   2000US-0229345P.
PR   05-SEP-2000;   2000US-0229509P.
PR   05-SEP-2000;   2000US-0229513P.
PR   06-SEP-2000;   2000US-0230437P.
PR   06-SEP-2000;   2000US-0230438P.
PR   08-SEP-2000;   2000US-0231242P.
PR   08-SEP-2000;   2000US-0231243P.
PR   08-SEP-2000;   2000US-0231244P.
PR   08-SEP-2000;   2000US-0231413P.
PR   08-SEP-2000;   2000US-0231414P.
PR   08-SEP-2000;   2000US-0232080P.
PR   08-SEP-2000;   2000US-0232081P.
PR   12-SEP-2000;   2000US-0231968P.
PR   14-SEP-2000;   2000US-0232397P.
PR   14-SEP-2000;   2000US-0232398P.
PR   14-SEP-2000;   2000US-0232399P.
PR   14-SEP-2000;   2000US-0232400P.
PR   14-SEP-2000;   2000US-0232401P.
PR   14-SEP-2000;   2000US-0233063P.
PR   14-SEP-2000;   2000US-0233064P.
PR   14-SEP-2000;   2000US-0233065P.
PR   21-SEP-2000;   2000US-0234223P.
PR   21-SEP-2000;   2000US-0234274P.
PR   25-SEP-2000;   2000US-0234937P.
PR   25-SEP-2000;   2000US-0234998P.
PR   26-SEP-2000;   2000US-0235484P.
PR   27-SEP-2000;   2000US-0235834P.
PR   27-SEP-2000;   2000US-0235836P.
PR   29-SEP-2000;   2000US-0236327P.
PR   29-SEP-2000;   2000US-0236367P.
PR   29-SEP-2000;   2000US-0236368P.
PR   29-SEP-2000;   2000US-0236369P.
PR   29-SEP-2000;   2000US-0236370P.
PR   02-OCT-2000;   2000US-0236802P.
PR   02-OCT-2000;   2000US-0237037P.
PR   02-OCT-2000;   2000US-0237038P.
PR   02-OCT-2000;   2000US-0237039P.
PR   02-OCT-2000;   2000US-0237040P.
PR   13-OCT-2000;   2000US-0239935P.
PR   13-OCT-2000;   2000US-0239937P.
PR   20-OCT-2000;   2000US-0240960P.
PR   20-OCT-2000;   2000US-0241221P.
PR   20-OCT-2000;   2000US-0241785P.
PR   20-OCT-2000;   2000US-0241786P.
PR   20-OCT-2000;   2000US-0241787P.
PR   20-OCT-2000;   2000US-0241808P.
PR   20-OCT-2000;   2000US-0241809P.
PR   20-OCT-2000;   2000US-0241826P.
PR   01-NOV-2000;   2000US-0244617P.
PR   08-NOV-2000;   2000US-0246474P.
PR   08-NOV-2000;   2000US-0246475P.
PR   08-NOV-2000;   2000US-0246476P.
PR   08-NOV-2000;   2000US-0246477P.

PR   08-NOV-2000;   2000US-0246478P.
PR   08-NOV-2000;   2000US-0246523P.
PR   08-NOV-2000;   2000US-0246524P.
PR   08-NOV-2000;   2000US-0246525P.
PR   08-NOV-2000;   2000US-0246526P.
PR   08-NOV-2000;   2000US-0246527P.
PR   08-NOV-2000;   2000US-0246528P.
PR   08-NOV-2000;   2000US-0246532P.
PR   08-NOV-2000;   2000US-0246609P.
PR   08-NOV-2000;   2000US-0246610P.
PR   08-NOV-2000;   2000US-0246611P.
PR   08-NOV-2000;   2000US-0246613P.
PR   17-NOV-2000;   2000US-0249207P.
PR   17-NOV-2000;   2000US-0249208P.
PR   17-NOV-2000;   2000US-0249209P.
PR   17-NOV-2000;   2000US-0249210P.
PR   17-NOV-2000;   2000US-0249211P.
PR   17-NOV-2000;   2000US-0249212P.
PR   17-NOV-2000;   2000US-0249213P.
PR   17-NOV-2000;   2000US-0249214P.
PR   17-NOV-2000;   2000US-0249215P.
PR   17-NOV-2000;   2000US-0249216P.
PR   17-NOV-2000;   2000US-0249217P.
PR   17-NOV-2000;   2000US-0249218P.
PR   17-NOV-2000;   2000US-0249244P.
PR   17-NOV-2000;   2000US-0249245P.
PR   17-NOV-2000;   2000US-0249264P.
PR   17-NOV-2000;   2000US-0249265P.
PR   17-NOV-2000;   2000US-0249297P.
PR   17-NOV-2000;   2000US-0249299P.
PR   17-NOV-2000;   2000US-0249300P.
PR   01-DEC-2000;   2000US-0250160P.
PR   01-DEC-2000;   2000US-0250391P.
PR   05-DEC-2000;   2000US-0251030P.
PR   05-DEC-2000;   2000US-0251988P.
PR   05-DEC-2000;   2000US-0257719P.
PR   06-DEC-2000;   2000US-0251479P.
PR   08-DEC-2000;   2000US-0251856P.
PR   08-DEC-2000;   2000US-0251868P.
PR   08-DEC-2000;   2000US-0251869P.
PR   08-DEC-2000;   2000US-0251989P.
PR   08-DEC-2000;   2000US-0251990P.
PR   11-DEC-2000;   2000US-0254097P.
PR   05-JAN-2001;   2001US-0259678P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA, Barash SC, Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
      useful for preventing, diagnosing and/or treating cancers and metastasis.
XX
PS   Disclosure; SEQ ID NO 25558; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting the
CC   nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention

```
XX
SQ   Sequence 22700 BP; 5667 A; 5046 C; 5162 G; 6825 T; 0 U; 0 Other;

Query Match        15.6%;   Score 31.4;   DB 4;   Length 22700;
Best Local Similarity  59.6%;   Pred. No. 14;
Matches   53;   Conservative    0;   Mismatches   36;   Indels    0;   Gaps    0;

Qy     68 GTTTAGCCTGTGTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAAAGGGGGATCTGA 127
          ||||||||| ||||||||||||||| ||| ||||||| ||||||||| | ||||||
Db   8161 GTATATATGGCTCCAGGGCTGTTACATGCCTACCCGTAAGTGGGACTAAAGGATCTTC 8220

Qy    128 TTATTGAGGTGTGGAAGGAATAAATAATC 156
          || |||| |||| |||  ||||
Db   8221 TTCAAAGGTGAGGATGTTGTGGCTACTC 8249


RESULT 7
AAL54072
ID   AAL54072 standard; DNA; 213456 BP.
XX
AC   AAL54072;
XX
DT   20-MAR-2003  (first entry)
XX
DE   Genomic DNA encoding a human transporter protein.
XX
KW   Cytostatic; human transporter protein; immune response; tissue typing;
KW   drug screening assay; pharmacogenomic analysis; placenta; nervous system;
KW   brain; hippocampus; transgenic animal; gene therapy; human; gene; ds.
XX
OS   Homo sapiens.
XX
FH   Key          Location/Qualifiers
FT   variation    replace(177,C)
FT                /*tag= u
FT                /standard_name= "Single nucleotide polymorphism"
FT   CDS          2189..21806
FT                /*tag= a
FT                /product= "Human transporter protein"
FT   exon         2189..2270
FT                /*tag= b
FT                /number= 1
FT   intron       2271..5227
FT                /*tag= c
FT                /number= 1
FT   variation    replace(2555,T)
FT                /*tag= v
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(2924,A)
FT                /*tag= w
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(3162,T)
FT                /*tag= x
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(3163,T)
FT                /*tag= y
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(4720,C)
FT                /*tag= z
FT                /standard_name= "Single nucleotide polymorphism"
FT   exon         5228..5365
FT                /*tag= d
FT                /number= 2
FT   intron       5366..158436
FT                /*tag= e
FT                /number= 2
FT   variation    replace(6289,A)
FT                /*tag= aa
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(6611,T)
FT                /*tag= ab
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(7861,G)
FT                /*tag= ac
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(8614,T)
FT                /*tag= ad
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(9249,G)
FT                /*tag= ae
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(10272,C)
FT                /*tag= af
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(10351,T)
FT                /*tag= ag
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(10444,A)
FT                /*tag= ah
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(10467,A)
FT                /*tag= ai
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(10785,T)
FT                /*tag= aj
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(10846,T)
FT                /*tag= ak
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(10847,G)
FT                /*tag= al
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(13842,G)
FT                /*tag= am
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(14103,G)
FT                /*tag= an
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(14128,C)
FT                /*tag= ao
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(15560,A)
FT                /*tag= ap
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(15736,C)
FT                /*tag= aq
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(15914,T)
FT                /*tag= ar
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(17099,G)
FT                /*tag= as
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(18139,A)
FT                /*tag= at
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(19303,T)
FT                /*tag= au
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(19501,A)
FT                /*tag= av
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(22852,A)
FT                /*tag= aw
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(23210,G)
FT                /*tag= ax
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(24040,A)
FT                /*tag= ay
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(24536,A)
FT                /*tag= az
FT                /standard_name= "Single nucleotide polymorphism"
FT   variation    replace(25125,A)
FT                /*tag= ba
FT                /standard_name= "Single nucleotide polymorphism"
```

```
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(25167,T)
FT                   /*tag= bb
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(25623,T)
FT                   /*tag= bc
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(27143,T)
FT                   /*tag= bd
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(27211,G)
FT                   /*tag= be
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(27383,A)
FT                   /*tag= bf
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(27553,T)
FT                   /*tag= bg
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(29141,G)
FT                   /*tag= bh
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(30549,T)
FT                   /*tag= bi
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(30984,T)
FT                   /*tag= bj
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(31266,A)
FT                   /*tag= bk
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(31425,C)
FT                   /*tag= bl
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(32198,G)
FT                   /*tag= bm
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(33129,C)
FT                   /*tag= bn
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(33292,G)
FT                   /*tag= bo
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(34124,T)
FT                   /*tag= bp
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(34668,A)
FT                   /*tag= bq
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(34807,T)
FT                   /*tag= br
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(36025,C)
FT                   /*tag= bs
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(36058,C)
FT                   /*tag= bt
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(36061,C)
FT                   /*tag= bu
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(36418,C)
FT                   /*tag= bv
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(36743,G)
FT                   /*tag= bw
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(36958,C)
FT                   /*tag= bx
FT   variation
FT                   /standard_name= "Single nucleotide polymorphism"
FT                   replace(37577,G)
FT                   /*tag= by
```

```
Query Match              15.5%;  Score 31.2;  DB 7;  Length 213456;
Best Local Similarity    52.3%;  Pred. No. 45;
Matches    69;  Conservative   0;  Mismatches  63;  Indels  0;  Gaps  0;

Qy      8 TTCTCCCTGAAGCCCTTCCCTGCCAACCAGCCAGCAACTTCCAAAATTCTGCCCGT 67
          |||||||| || ||| |||||| ||| || ||||| | ||||||  ||||||||
Db  64946 TTCTCCCTACCAGTATTCGTAGAACCATATGCCAGTTAAACTCCACCATTTTGCCTCT 65005

Qy     68 GTTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAAGGGGATCTGA 127
          ||  || | | | ||   ||||| ||| ||  || |||||||  |||||||| |||| ||
Db  65006 TAGTGTGCATGCTTGAGCTCATTCACCCAGTTTCTGAGATATTGGGAAAATGCGATCACC 65065

Qy    128 TTATTGAGGTGT 139
          || ||||||| ||
Db  65066 AGTTTCAGGTTT 65077


RESULT 8
ABZ11734
ID   ABZ11734 standard; cDNA; 1830 BP.
XX
AC   ABZ11734;
XX
DT   20-JAN-2003  (first entry)
DE   Human polynucleotide SEQ ID NO 616.
XX
KW   Human; genome mapping; gene therapy; food supplement; virus; fungus;
KW   cell-proliferative disorder; neurodegenerative disease; bacterial;
KW   Parkinson's disease; Alzheimer's disease; autoimmune disease;
KW   multiple sclerosis; diabetes; genetic disorder; wound; burn; infection;
KW   arthritis; cytostatic; immunomodulator; nootropic; neuroprotective;
KW   antiparkinsonian; antidiabetic; immunosuppressive; dermatological;
KW   haemostatic; vulnerary; fungicide; antibacterial; virucide; protozoacide;
KW   antiarthritic; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO20270539-A2.
XX
PD   12-SEP-2002.
XX
PF   05-MAR-2002; 2002WO-US005095.
XX
PR   05-MAR-2001; 2001US-00799451.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Zhou P,  Goodrich RW,  Asundi V,  Zhang J,  Zhao QA,  Ren F;
PI   Xue AJ,  Yang Y,  Ma Y,  Yamazaki V,  Chen R,  Wang Z,  Ghosh M;
PI   Wehrman T,  Wang J,  Wang D,  Drmanac RT;
XX
DR   WPI; 2002-759812/82.
DR   P-PSDB; ABP69517.
XX
PT   New polynucleotides comprising sequences assembled from expressed
PT   sequence tags (ESTs), useful for treating cell-proliferative,
PT   neurodegenerative, autoimmune, genetic, myeloid or lymphoid, or platelet
PT   or coagulation disorders.
XX
PS   Claim 1; SEQ ID NO 616; 1012pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated polynucleotide (I) comprising a
CC   nucleotide sequence selected from any of 948 sequences (ABZ11119-
CC   ABZ12066) or their mature protein coding portion, active domain coding
CC   protein or complementary sequences. The polynucleotides are useful for
CC   identifying expressed genes or for physical mapping of human genome. The
CC   encoded polypeptides (ABP68902-ABP69849) are useful as molecular weight
CC   markers, as a food supplement, for generating antibodies, in medical
CC   imaging, screening and diagnostic assays and for treating cell-
CC   proliferative disorders (cancer), neurodegenerative diseases (Parkinson's
CC   or Alzheimer's disease), autoimmune diseases (multiple sclerosis,
CC   diabetes, lupus) genetic disorders, myeloid or lymphoid disorders,
```

```
CC   platelet or coagulation disorders, wound, burns, incision, ulcers, liver
CC   or lung fibrosis, infections (bacterial, viral, fungal, parasitic),
CC   arthritis, etc. Note: The sequence data for this patent did not form part
CC   of the printed specification, but was obtained in electronic format
CC   directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1830 BP; 643 A; 283 C; 357 G; 547 T; 0 U; 0 Other;

Query Match             15.4%;  Score 31;  DB 6;  Length 1830;
Best Local Similarity   56.3%;  Pred. No. 6.4;
Matches   58;  Conservative   0;  Mismatches  45;  Indels   0;  Gaps   0;

Qy         76  TTGTTAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAAGGGGATCTGATTATTGAG 135
               || ||||||| || ||||||||  || |||||||||||||| | |||| |||| ||||
Db        640  TTTTTAAGGGGGTTCAGATCCTGGAAAGAAAATGCCAATAGTGACTATGAGAATGTAG  699

Qy        136  GTGTGGAAGGAATAAATAATCAGTCCACAAATAAACAAACTGT 178
               |||||    |||  |||||||||||||||||||  |||||  ||
Db        700  AAGTGGCATCTGAAAACAATTGGTACACAAGTAAACTCGGCAGT 742
```

```
RESULT 9
AAD53223_3
Continuation (4 of 5) of AAD53223 from base 300001 (Human chromosome 3 p-arm breakpoint
WP Sequence split into 5 fragments  LOCUS AAD53223 Accession Aad53223
WP  Fragment Name          Begin       End
WP  AAD53223_0                 1    110000
WP  AAD53223_1            100001    210000
WP  AAD53223_2            200001    310000
WP  AAD53223_3            300001    410000
WP  AAD53223_4            400001    487980

Query Match             15.4%;  Score 31;  DB 7;  Length 110000;
Best Local Similarity   53.8%;  Pred. No. 39;
Matches   64;  Conservative   0;  Mismatches  55;  Indels   0;  Gaps   0;

Qy         57  ATTCTGCCCGTGTTGTTAGCCTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAA 116
               |||| ||    |||||| |||| || |||||| ||| ||||||||||||||||||||||||
Db      96672  AGTTTGGCTAATTTGAATCCTGTTCAAGGGTTTCAGGGTTTTCCCTTACTTGCTGCCAAAATTGATT 96731

Qy        117  AGGGGATCTGATTATTGAGGTGTGGAAGAATAATAATCAGTCCACAAATAAACAAAC 175
               || |||||||||||||||||  || ||| |||||||||||||| ||||||||||||||
Db      96732  TGTGGATTTTCTCCATGAAGTTAGAGGGCCAGAAAATAATGAGGGCCAGGAAGAATGAAC 96790
```

```
RESULT 10
ABX41793/c
ID   ABX41793 standard; cDNA; 360 BP.
XX
AC   ABX41793;
XX
DT   20-FEB-2003  (first entry)
XX
DE   Bovine EST associated with lactation/muscle/fat deposition #6958.
XX
KW   Bovine; ss; EST; expressed sequence tag; lactation; LMFD;
KW   muscle deposition; fat deposition; genome mapping; gene identification;
KW   gene analysis; cattle breeding.
XX
OS   Bos Taurus.
XX
PN   US2002137139-A1.
XX
PD   26-SEP-2002.
XX
PF   24-SEP-2001; 2001US-00960352.
XX
PR   12-JAN-1999;   99US-0115707P.
PR   11-JAN-2000; 2000US-0480902.
XX
PA   (BYAT/) BYATT J C.
PA   (MATH/) MATHIALAGAN N.
PA   (TAON/) TAO N.
```

```
PA   (WARR/) WARREN W C.
XX
PI   Byatt JC, Mathialagan N, Tao N, Warren WC;
XX
DR   WPI; 2003-110599/10.
XX
PT   New nucleic acid associated with lactation, and muscle and fat
PT   deposition, useful for genome mapping, gene identification and analysis,
PT   cattle breeding, or for genetically improving cattle.
XX
PS   Claim 2; SEQ ID NO 6958; 245pp; English.
XX
CC   The invention relates to a purified nucleic acid molecule associated with
CC   lactation or muscle and fat deposition (designated LMFD), derived from
CC   cattle, and the LMFD nucleic acid can specifically hybridise to a second
CC   nucleic acid molecule comprising any of 15112 nucleotide sequences,
CC   appearing as ABX34836-ABX49947, or complements of them. Also included are
CC   : (1) a transfomed cell having a nucleic acid comprising an LMFD nucleic
CC   acid linked to a promoter and a 3'-non- translated sequence that
CC   functions in the cell to cause termination of transcription and addition
CC   of polyadenylated ribonucleotides to a 3' end of the mRNA molecule; and
CC   (2) determining a level or pattern of a molecule in a bovine cell or
CC   tissue comprising: (a) incubating a marker nucleic acid (comprising any
CC   of the 15112 nucleic acid sequences or its complement or fragment) with a
CC   complementary nucleic acid molecule obtained from the bovine cell or
CC   tissue, where hybridisation between the marker nucleic acid and the
CC   complementary nucleic acid permits the detection of the molecule; and (b)
CC   detecting the level or pattern of the complementary nucleic acid, where
CC   the detection of the complementary nucleic acid is predictive of the
CC   level or pattern of the molecule. The LMFD nucleic acid is used for
CC   determining a level or pattern of a molecule in a bovine cell or tissue.
CC   It is useful for genome mapping, gene identification and analysis, cattle
CC   breeding, preparation of constructs for use in cattle gene expression, or
CC   for genetically improving cattle. The present sequence is one of the
CC   15112 bovine LMFD EST (expressed sequence tag) nucleic acids. Note: The
CC   present sequence was not shown in the specification but was obtained in
CC   electronic format from the USPTO web site:
CC   seqdata.uspto.gov/sequence.html?DocID=20020137139
XX
SQ   Sequence 360 BP; 118 A; 69 C; 47 G; 126 T; 0 U; 0 Other;

Query Match             15.3%;  Score 30.8;  DB 7;  Length 360;
Best Local Similarity   55.7%;  Pred. No. 3.7;
Matches   59;  Conservative   0;  Mismatches  47;  Indels   0;  Gaps   0;

Qy         54  AAAATTCTGCCCGTGTTTAGCCTTGTTAAAGGGGGTGTCTCACTCCTTCAGGGAAAGTGG 113
               |||||| |||| |||||    || |||||||| |||||||||||||||||||||||||||
Db        347  AAAATTTTTATTCTCATTAAACTTTTTTTAAATGGGTATAAAATTTTGTGACAGATTTTTG 288

Qy        114  AAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAAATCAGT 159
               |||||||||||||||||||| | |||||||  |||||||| |  ||||
Db        287  AAAAGTTGTTTTTTCATTAAAAGTACTGATTTTAAAAACTAAATATT 242
```

```
RESULT 11
ABS30387
ID   ABS30387 standard; DNA; 491 BP.
XX
AC   ABS30387;
XX
DT   25-FEB-2003  (first entry)
XX
DE   Human liver single exon probe, SEQ ID No 5377.
XX
KW   Human; single exon nucleic acid probe; liver; cirrhosis;
KW   hyperlipoproteinaemia; hyperlipidaemia; hypercholesterolaemia;
KW   coronary heart disease; ss.
XX
OS   Homo sapiens.
XX
PN   WO20015273-A2.
XX
PD   09-AUG-2001.
```

XX
PF 30-JAN-2001; 2001WO-US000664.
XX
PR 04-FEB-2000; 2000US-0180312P.
PR 26-MAY-2000; 2000US-0207456P.
PR 30-JUN-2000; 2000US-06608408.
PR 03-AUG-2000; 2000US-00632366.
PR 21-SEP-2000; 2000US-0234687P.
PR 27-SEP-2000; 2000US-0236359P.
PR 04-OCT-2000; 2000GB-00024263.
XX
PA (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI Penn SG, Hanzel DK, Chen W, Rank DR;
XX
DR WPI; 2001-488898/53.
XX
PT Human genome-derived single exon nucleic acid probes useful for analyzing
PT gene expression in human adult liver.
XX
PS Claim 1; SEQ ID NO 5377; 658pp; English.
XX
CC The invention relates to a single exon nucleic acid probe (SENP) (I) for
CC measuring human gene expression in a sample derived from human adult
CC liver, comprising one of 13109 defined nucleotide sequences given in the
CC specification (or complements/ fragments). The probe hybridises at high
CC stringency to a nucleic acid molecule expressed in the human adult liver.
CC (I) may be used for predicting, measuring and displaying gene expression
CC in samples derived from human adult liver. The genes identified may be
CC involved in genetic liver diseases such as cirrhosis,
CC hyperlipoproteinaemia, hyperlipidaemia and hypercholesterolaemia which is
CC associated with coronary heart disease. ABS25011-ABS51005 represent human
CC liver single exon nucleic acid probes of the invention. Note: The
CC sequence information for this patent does not appear in the printed
CC specification but was obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 491 BP; 135 A; 110 C; 144 G; 102 T; 0 U; 0 Other;

Query Match   15.1%; Score 30.4; DB 4; Length 491;
Best Local Similarity 71.4%; Pred. No. 5.8;
Matches   40; Conservative   0; Mismatches   16; Indels   0; Gaps   0;

Qy   119 GGGATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTCCACAAATAAACAAA 174
            |||||||||||| ||||| ||||| |||| ||||||| || | ||| |||
Db   1 GGGATCTGATTAATGAGGGATTGTAGGGAGTAAAGAGACTCTAAAAATATAGAAA 56

RESULT 12
AAX56241
ID AAX56241 standard; DNA; 3574 BP.
XX
AC AAX56241;
XX
DT 16-JUL-1999 (first entry)
XX
DE Mouse JMY nucleotide sequence.
XX
KW Mouse; JMY; co-activator; p300/CBP; cell cycle; ss.
XX
OS Mus sp.
XX
PN WO9920752-A1.
XX
PD 29-APR-1999.
XX
PF 21-OCT-1998; 98WO-GB003152.
XX
PR 21-OCT-1997; 97GB-00022238.
PR 20-AUG-1998; 98GB-0018235.
XX
PA (UNIU) UNIV GLASGOW.
XX

PI La Thangue NB, Shikama N;
XX
DR WPI; 1999-302738/25.
DR P-PSDB; AAY09513.
XX
PT JMY, a co-activator for p300/CBP.
XX
PS Claim 8; Page 72-74; 81pp; English.
XX
CC The present sequence encodes mouse JMY, which is a co-activator of
CC p300/CBP. The JMY polynucleotide (I), its complements or fragments (such
CC as primers) can be used to detect the presence of a JMY gene in a sample.
CC Modulators of JMY can be used to regulate the cell cycle. JMY antibodies
CC can be used to detect these modulators
XX
SQ Sequence 3574 BP; 954 A; 905 C; 1029 G; 686 T; 0 U; 0 Other;

Query Match   15.1%; Score 30.4; DB 2; Length 3574;
Best Local Similarity 52.3%; Pred. No. 14;
Matches   67; Conservative   0; Mismatches   61; Indels   0; Gaps   0;

Qy   9 TCTCCCTGAAGCCCTCTTCCGTGCCCAACCAGACCAGCAACTTCCAAAATTCTGCCCGTG 68
            ||||||| || || |||||||| ||| ||| || ||| | |||||| |||||
Db   2520 TCTTCCTCCACACTTCCCCTCCCCCTCCCACCCCCACCCCCACCCCCACCCCCTGCC 2579
Qy   69 TTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAGGTGGGAAAAGGGGATCTGAT 128
            || || ||| ||| || | ||||| |||| ||| |||| || |
Db   2580 TGTTCAAAGGACAATGGGGCCTCCACCACTGCAGAGACACTGGAGAAAGATGCACTTAG 2639
Qy   129 TATTGAGG 136
            ||||||
Db   2640 GACGGAGG 2647

RESULT 13
AAA72296/c
ID AAA72296/c standard; DNA; 1466 BP.
XX
AC AAA72296;
XX
DT 06-DEC-2000 (first entry)
XX
DE Drosophila odorant receptor DOR 47E.1 genomic DNA.
XX
KW Odorant receptor; Drosophila; olfactory receptor;
KW G protein-coupled receptor; GPCR superfamily; transgenic insect;
KW insect behaviour modification; pest control; pollinator attraction;
KW biosensor; odour detection; odour identification; apiculture; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200043410-A2.
XX
PD 27-JUL-2000.
XX
PF 25-JAN-2000; 2000WO-US001823.
XX
PR 25-JAN-1999; 99US-0117132P.
XX
PA (UYYA) UNIV YALE.
XX
PI Carlson JR, Kim J, Clyne PJ, Warr CG;
XX
DR WPI; 2000-543246/49.
DR P-PSDB; AAB20912.
XX
PT New nucleic acid encoding a Drosophila olfactory receptor, useful for
PT identifying modulating agents.
XX
PS Example 1; Page 300-301; 303pp; English.
XX
CC Sequences AAA72285-A72298 represent genomic DNA encoding the Drosophila
CC melanogaster odorant receptors AAB20901-B20914. Drosophila odorant

CC   receptor proteins (AAB20901-B20949) function as olfactory receptors, and
CC   are thought to be members of the G protein-coupled receptor (GPCR)
CC   superfamily, which is characterised by the presence of 7 transmembrane
CC   helices. Nucleic acids encoding the Drosophila odorant receptors may be
CC   used to generate expression constructs, host cells containing such
CC   constructs, and transgenic insects. Cells which express the odorant
CC   receptor genes may be used in methods to identify agents which modulate
CC   expression of these genes, and in methods to identify receptor binding
CC   partners. The Drosophila odorant receptor nucleic acids may also be used
CC   to identify corresponding genes in other insects, such as those which
CC   damage crops or transmit disease. The odorant receptor proteins may be
CC   used to identify agents which modulate their activity, to identify
CC   binding partners, as antigens to raise antibodies, and in methods to
CC   modify insect behaviour. The proteins may be also be used in methods of
CC   behaviour modification. Such methods may be used to study or modify
CC   insect behaviour in response to odorants such as pheromones. Modification
CC   of insect behaviour has a wide range of applications, such as in pest
CC   control (e.g., by disrupting the feeding or mating behaviours of pest
CC   species), or for enhancing plant pollination (by attracting pollinator
CC   species). Odorant receptor proteins and/or nucleotides may also be used
CC   to identify appetite suppressants, to trap odours of a specific type, as
CC   biosensors for the detection of explosives, drugs, perfumes or
CC   pollutants, and in apiculture to modify the behaviour of bees, for
CC   example, to increase the production of royal jelly
XX
SQ   Sequence 1466 BP; 352 A; 341 C; 324 G; 449 T; 0 U; 0 Other;

Query Match   15.0%; Score 30.2; DB 3; Length 1466;
Best Local Similarity   55.1%;   Pred. No. 11;
Matches   59; Conservative   0; Mismatches   48; Indels   0; Gaps   0;

Qy   48 ACTTCCAAAATTCTGCCCGTGTTGTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAA 107
              |||| ||||| | || | |||||||| ||| |||||||||| ||| | ||| |||||||
Db   141 ACTTGCAGAAAACTGTCCATGTCGACAAGGGTTAATTCGGCCTCACAAGTCAAGTGT 82

Qy   108 AGTGGGAAAAGGGGATCTGATTATTGAGGTGTGGAAGAATAAATAA 154
              |||||||||| |||| ||||||||||||||||| |||||||||
Db   81 GTTCAGTTCAGGGGTTTTAAACCTTTCAATTGAGTGAGGAATAAATGA 35

RESULT 14
AAA94812/c
ID   AAA94812 standard; DNA; 2075 BP.
XX
AC   AAA94812;
XX
DT   23-FEB-2001 (first entry)
XX
DE   Drosophila melanogaster odorant receptor DOR24 coding sequence.
XX
KW   Odorant receptor; fruit fly; DOR24; odour recognition; pest control; ds.
XX
OS   Drosophila melanogaster.
XX
FH   Key   Location/Qualifiers
FT   CDS   1..2075
FT   /*tag= a
FT   /product= "DOR24"
FT   intron   187..739
FT   /*tag= b
FT   /number= 1
FT   /cons_splice= (5'site:NO,3'site:NO)
FT   intron   1037..1338
FT   /*tag= c
FT   /number= 2
FT   /cons_splice= (5'site:NO,3'site:NO)
FT   intron   1954..2024
FT   /*tag= d
FT   /number= 3
FT   /cons_splice= (5'site:NO,3'site:YES)
XX
PN   WO200050566-A2.

XX
PD   31-AUG-2000.
XX
PF   25-FEB-2000; 2000WO-US004995.
XX
PR   25-FEB-1999; 99US-00257706.
XX
PA   (UYCO) UNIV COLUMBIA NEW YORK.
XX
PI   Vosshall LB, Amrein HO, Axel R;
XX
DR   WPI; 2000-572081/53.
DR   P-PSDB; AAB26412.
XX
PT   Novel nucleic acid encoding an insect odorant receptor, for identifying
PT   modulator compounds that are useful in controlling pest population.
XX
PS   Claim 4; Fig 8; 176pp; English.
XX
CC   The present sequence is the coding sequence for the Drosophila
CC   melanogaster odorant receptor DOR24. It was isolated using a differential
CC   cloning strategy. The odorant genes and proteins are useful as they aid
CC   in the study of the olfactory organ in mammals, as well as aiding the
CC   understanding of the link between odour recognition and behaviour in
CC   insects. They also enable the identification of compounds capable of
CC   activating and inhibiting the receptors, allow the control of pest
CC   populations via the use of alarm odour ligands and via the use of ligands
CC   which interfere with the interaction between odorant ligands and
CC   receptors associated with fertility
XX
SQ   Sequence 2075 BP; 548 A; 460 C; 447 G; 620 T; 0 U; 0 Other;

Query Match   15.0%; Score 30.2; DB 3; Length 2075;
Best Local Similarity   55.1%;   Pred. No. 13;
Matches   59; Conservative   0; Mismatches   48; Indels   0; Gaps   0;

Qy   48 ACTTCCAAAATTCTGCCCGTGTTGTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAA 107
              |||| ||||| | || | |||||||| ||| |||||||||| ||| | ||| |||||||
Db   750 ACTTGCAGAAAACTGTCCATGTCGACAAGGGTTAATTCGGCCTCACAAGTCAAGTGT 691

Qy   108 AGTGGGAAAAGGGGATCTGATTATTGAGGTGTGGAAGAATAAATAA 154
              |||||||||| |||| ||||||||||||||||| |||||||||
Db   690 GTTCAGTTCAGGGGTTTTAAACCTTTCAATTGAGTGAGGAATAAATGA 644

RESULT 15
ABL22180
ID   ABL22180 standard; DNA; 3548 BP.
XX
AC   ABL22180;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster genomic polynucleotide SEQ ID NO 18013.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical; gene; ds.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX

DR   WPI; 2001-656860/75.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Claim 1; SEQ ID NO 18013; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3548 BP; 1030 A; 784 C; 745 G; 989 T; 0 U; 0 Other;

Query Match   15.0%; Score 30.2; DB 4; Length 3548;
Best Local Similarity   55.1%; Pred. No. 16;
Matches   59; Conservative   0; Mismatches   48; Indels   0; Gaps   0;

Qy   48 ACTTCCAAAATTCGCCCGTGTTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAA 107
             || ||| || ||||| ||||| || || ||| ||| || | |||| | |||||
Db   2519 ACTTGCAGAAAACTGTCCATGTCGACACAGGGTTAATTCGGCCTCACACTAAGTCAAGTGT 2578

Qy   108 AGTGGGAAAAAGGGGATCTCGATTATTGAGGTGTCGGAAGGAATAAATAA 154
             || |||| | || | || ||| ||| |||| | ||||| |||||
Db   2579 GTTCAGTTCAGGGGTTTTAAACCTTTCAATTGAGTAGGAATAAATGA 2625

Search completed: March 25, 2004, 10:25:28
Job time : 212.591 secs

```
                    GenCore version 5.1.6
        Copyright (c) 1993 - 2004 Compgen Ltd.

OM nucleic - nucleic search, using sw model

Run on:   March 25, 2004, 09:55:14 ; Search time 39.5358 Seconds
                                    (without alignments)
                              2821.370 Million cell updates/sec

Title:          US-09-963-285-1_COPY_223_423
Perfect score:  201
Sequence:       1 acaaatgtctccctgagc.......ggattcctagaggaggaggag 201

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters:     1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   Issued Patents NA:*
         1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
         2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
         3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
         4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
         5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
         6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
          score greater than or equal to the score of the result being printed,
          and is derived by analysis of the total score distribution.

                              SUMMARIES
```

Result No.	Score	% Query Match	DB	Length	ID	Description
1	31.4	15.6	4	296	US-09-313-294A-330	Sequence 330, App
2	30.4	15.1	4	3574	US-09-311-236-1	Sequence 1, Appli
3 c	30.2	15.0	4	1466	US-09-491-577-110	Sequence 110, App
4	29.8	14.8	4	274	US-09-313-294A-4645	Sequence 4645, Ap
5	29.4	14.6	4	1191	US-09-252-991A-14425	Sequence 14425, A
6 c	29.4	14.6	4	640681	US-09-790-988-1	Sequence 1, Appli
7 c	29.2	14.5	1	513	US-08-031-143B-70	Sequence 70, Appl
8	29.2	14.5	4	1664976	US-08-916-421B-1	Sequence 1, Appli
9	28.6	14.2	4	2920	US-08-976-259-10	Sequence 10, Appl
10	28.2	14.0	4	3436	US-09-221-017B-335	Sequence 335, App
11 c	28.2	14.0	4	1664976	US-08-916-421B-1	Sequence 1, Appli
12	28	13.9	4	239	US-09-439-313-239	Sequence 239, App
13	28	13.9	4	239	US-09-352-616A-239	Sequence 239, App
14	28	13.9	4	239	US-09-232-149A-239	Sequence 239, App
15	28	13.9	4	239	US-09-159-812-239	Sequence 239, App
16	28	13.9	4	239	US-09-636-215-239	Sequence 239, App
17	28	13.9	4	239	US-09-685-166A-239	Sequence 239, App
18	28	13.9	4	239	US-09-688-489-239	Sequence 239, App
19 c	27.8	13.8	4	5829	US-08-809-254A-4	Sequence 4, Appli
20 c	27.8	13.8	4	41100	US-09-755-665-46	Sequence 46, Appl
21 c	27.2	13.5	4	832	US-09-621-976-2813	Sequence 2813, Ap
22	27.2	13.5	2	2402	US-09-023-655-1317	Sequence 1317, Ap
23	27.2	13.5	4	2430	US-08-488-199-3	Sequence 3, Appli
24	27.2	13.5	4	4797	US-09-643-597-134	Sequence 134, App
25	27.2	13.5	4	4797	US-09-480-884A-134	Sequence 134, App
26	27.2	13.5	4	4797	US-09-542-615A-134	Sequence 134, App
27	27.2	13.5	4	4797	US-09-606-421B-134	Sequence 134, App

28	27.2	13.5	4	4797	US-09-221-107-134	Sequence 134, App
29 c	27	13.4	4	400	US-08-956-171E-3680	Sequence 3680, Ap
30	27	13.4	4	621	US-09-107-532A-1542	Sequence 1542, Ap
31 c	26.8	13.3	4	2628	US-09-294-531B-5	Sequence 5, Appli
32 c	26.8	13.3	2	7218	US-08-232-463-14	Sequence 14, Appl
33	26.8	13.3	4	35081	US-08-752-760A-1	Sequence 1, Appli
34	26.6	13.2	4	468	US-09-543-681A-1682	Sequence 1682, Ap
35	26.6	13.2	4	1089	US-09-543-681A-895	Sequence 895, App
36 c	26.6	13.2	3	2119	US-09-018-615-28	Sequence 28, Appl
37 c	26.6	13.2	4	2119	US-09-912-962-28	Sequence 28, Appl
38	26.6	13.2	4	2119	US-09-706-197-3	Sequence 3, Appli
39	26.6	13.2	4	2712	US-09-976-594-374	Sequence 374, App
40 c	26.4	13.1	4	400	US-08-956-171E-4012	Sequence 4012, Ap
41	26.4	13.1	4	1005	US-09-107-532A-2682	Sequence 2682, Ap
42	26.4	13.1	4	1249	US-08-956-171E-389	Sequence 389, App
43 c	26.2	13.0	4	315	US-08-956-171E-4413	Sequence 4413, Ap
44	26.2	13.0	4	543	US-09-134-000C-394	Sequence 394, App
45 c	26.2	13.0	2	1408	US-08-440-845D-11	Sequence 11, Appl

```
                              ALIGNMENTS

RESULT 1
US-09-313-294A-330
; Sequence 330, Application US/09313294A
; Patent No. 6476212
; GENERAL INFORMATION:
;  APPLICANT: Lalgudi, Raghunath V.
;  APPLICANT: Ito, Laura Y.
;  APPLICANT: Sherman, Bradley K.
;  TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;  FILE REFERENCE: PL-0017 US
;  CURRENT APPLICATION NUMBER: US/09/313,294A
;  CURRENT FILING DATE: 1999-05-14
;  NUMBER OF SEQ ID NOS: 7600
;  SOFTWARE: PERL Program
; SEQ ID NO 330
;  LENGTH: 296
;  TYPE: DNA
;  ORGANISM: Zea mays
;  FEATURE:
;  NAME/KEY: misc_feature
;  OTHER INFORMATION: Incyte ID No. 6476212 700549010H1
;  NAME/KEY: unsure
;  LOCATION: 22, 64, 167
;  OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-330

Query Match       15.6%;      Score 31.4;   DB 4;    Length 296;
Best Local Similarity   55.7%;   Pred. No. 0.065;
Matches    59;  Conservative   0;  Mismatches  47;  Indels   0;  Gaps    0;

Qy      91 CTCACTCCTTCAGGGAAAGTGGGAAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAA 150
           |||||| ||||||   |||||||   |||| |||| |  ||||| | |||||| |||||
Db      76 CTTACATACACCGTCAGGTCGGCAGGGAGGAGCAGGTAATACTGGTGTTGCTGTCATGC 135

Qy     151 ATATCAGTCCACAAATAAACAAACTGTCCGGGATTCCTAGAGGGA 196
           ||| ||    ||| |   |||||    |||||||||||||||
Db     136 TTTATGAGCCCAGATATAAGCACAGTGTCAGTGTCAGNAGACTAGAAAGGGA 181

RESULT 2
US-09-311-236-1
; Sequence 1, Application US/09311236
; Patent No. 6376215
; GENERAL INFORMATION:
;  APPLICANT: LaThangue, Nicholas B
;  APPLICANT: Shikama, No. 637621Siko
;  APPLICANT: The University Court of the University of Glasgow
;  TITLE OF INVENTION: JMY, a co-activator for p300/CBP, nucleic acid encoding
;  TITLE OF INVENTION: JMY and uses thereof
;  FILE REFERNCE: AHB/LP5731500
```

```
Qy    48  ACTTCCAAAATTCTGCCCGTGTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAA  107
              ||||  ||  ||||||||||                        ||||  |||  |||
Db   141  ACTTGCAGAAACTGTCCATGTCGACAAGGGTTAATTCGGCCTCACACTAAGTCAAGTGT   82

Qy   108  AGTGGGAAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAAATAA  154
              |||||||||||                 |||||  ||||||||||||
Db    81  GTTCAGTTCAGGGGTTTTAACCTTTCAATTGAGTAGGAATAAAATGA   35
```

RESULT 4
US-09-313-294A-4645
; Sequence 4645, Application US/0913294A
; Patent No. 6476212
; GENERAL INFORMATION:
;   APPLICANT: Lalgudi, Raghunath V.
;   APPLICANT: Ito, Laura Y.
;   APPLICANT: Sherman, Bradley K.
;   TITLE OF INVENTION: POLYNUCLEOTIDES AND POLYPEPTIDES DERIVED FROM CORN EAR
;   FILE REFERENCE: PL-0017 US
;   CURRENT APPLICATION NUMBER: US/09/313,294A
;   CURRENT FILING DATE: 1999-05-14
;   NUMBER OF SEQ ID NOS: 7600
;   SOFTWARE: PERL Program
; SEQ ID NO 4645
;   LENGTH: 274
;   TYPE: DNA
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: misc feature
;   OTHER INFORMATION: Incyte ID No. 6476212 700348780H1
;   NAME/KEY: unsure
;   LOCATION: 70,129
;   OTHER INFORMATION: a, t, c, g, or other
US-09-313-294A-4645

Query Match   14.8%;   Score 29.8; DB 4; Length 274;
Best Local Similarity   54.2%;   Pred. No. 0.25;
Matches   58; Conservative   0; Mismatches   49; Indels   0; Gaps   0;

```
Qy    91  CTCACTCCTTCAGGGAAAGTGGGAAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAA  150
              |||||||  ||  |||||  ||||  |||  ||  ||||||||  ||  |||||  |||
Db    65  CTTACNTACACCGGTCAGGTCGGACAGGGAGCAGGTAATACTGGTGTTGCTGTCATGC   124

Qy   151  ATAATCAGTCACAAATAAACAAACTGTCCCGGATCCTACGAGGGAA  197
              ||||  |||  ||||||||||  ||||||||||  ||||||||||
Db   125  TTTANGACCCCAGATATAAGTACAGTCAGCAGACTAGAAAGGGAA  171
```

RESULT 5
US-09-252-991A-14425
; Sequence 14425, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
;   APPLICANT: Marc J. Rubenfield et al.
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
;   TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
;   FILE REFERENCE: 107196.136
;   CURRENT APPLICATION NUMBER: US/09/252,991A
;   CURRENT FILING DATE: 1999-02-18
;   PRIOR APPLICATION NUMBER: US 60/074,788
;   PRIOR FILING DATE: 1998-02-18
;   PRIOR APPLICATION NUMBER: US 60/094,190
;   PRIOR FILING DATE: 1998-07-27
;   NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 14425
;   LENGTH: 1191
;   TYPE: DNA
;   ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-14425

Query Match   14.6%;   Score 29.4; DB 4; Length 1191;
Best Local Similarity   58.6%;   Pred. No. 0.82;
Matches   51; Conservative   0; Mismatches   36; Indels   0; Gaps   0;
```

```
; CURRENT APPLICATION NUMBER: US/09/311,236
; CURRENT FILING DATE: 1999-05-13
; EARLIER APPLICATION NUMBER: GB 9722238.4
; EARLIER FILING DATE: 1997-10-21
; EARLIER APPLICATION NUMBER: GB 9818235.5
; EARLIER FILING DATE: 1998-08-20
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
;   LENGTH: 3574
;   TYPE: DNA
;   ORGANISM: Mus sp.
US-09-311-236-1

Query Match               15.1%;   Score 30.4;  DB 4;  Length 3574;
Best Local Similarity     52.3%;   Pred. No. 0.66;
Matches       67;  Conservative   0;  Mismatches   61;  Indels    0;  Gaps    0;
```

```
Qy     9  ACTCCCTGAAGCCCTCTTCCCTGCCAACCAGACCAACTTCCAAAATTCTGCCCGTG   68
             ||||  |||||  |||  ||  |||||||||  |||||||||||||||||||||||
Db  2520  TCTTCCTCCAACACCTCCCCTCTCCCCACCCCCCCACCTCCACCCCCACCCCTGCC  2579

Qy    69  TTTAGCCCTTGTTAAAGGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAAGGGGATCTGAT  128
             |||||  ||  ||||||  |||  |||  ||||||||  |||||||||||
Db  2580  TGTTGCAAAGGACAATGGGCCCTCCACCACTGCAGAGACACTGGAGAAAGATGCACTTAG  2639

Qy   129  TATTGAGG  136
             ||||||||
Db  2640  GACGGAGG  2647
```

RESULT 3
US-09-491-577-110/c
; Sequence 110, Application US/09491577
; Patent No. 6610511
; GENERAL INFORMATION:
;   APPLICANT: Yale University
;   APPLICANT: Carlson, John R.
;   APPLICANT: Kim, Hunhyong
;   APPLICANT: Clyne, Peter J.
;   APPLICANT: Warr, Coral G.
;   TITLE OF INVENTION: No. 6610511el Family of Odorant Receptor Genes in Drosophila
;   FILE REFERENCE: 44574-5061-US
;   CURRENT APPLICATION NUMBER: US/09/491,577
;   CURRENT FILING DATE: 2000-01-25
;   EARLIER APPLICATION NUMBER: US 60/117,132
;   EARLIER FILING DATE: 1999-01-25
;   NUMBER OF SEQ ID NOS: 112
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 110
;   LENGTH: 1466
;   TYPE: DNA
;   ORGANISM: Drosophila melanogaster
;   FEATURE:
;   OTHER INFORMATION: DOR 47E.1, NCBI Accession No. 6610511 AF156880
;   FEATURE:
;   NAME/KEY: intron.
;   LOCATION: (1)..(121)
;   FEATURE:
;   NAME/KEY: intron.
;   LOCATION: (423)..(478)
;   FEATURE:
;   NAME/KEY: intron.
;   LOCATION: (669)..(728)
;   FEATURE:
;   NAME/KEY: intron
;   LOCATION: (1345)..(1415)
US-09-491-577-110

Query Match   15.0%;   Score 30.2; DB 4; Length 1466;
Best Local Similarity   55.1%;   Pred. No. 0.47;
Matches   59; Conservative   0; Mismatches   48; Indels   0; Gaps   0;
```

```
Qy    20  CCCCTTCCTGCCCAACCAGACCAGCAACTTCCAAAATTCTGCCCGTGTTAGCCTTGT   79
          ||||| ||| |||| |||||||
Db   898  CTCTTGGCTGCACAATGACTCCAGCCTGCTCCAATTCACTGCACCAATTCAGCCCGGA  957

Qy    80  TAAAGGGGTGTCTCACTCCTTCAGGGA  106
          || ||| ||||
Db   958  TCAAGTGGTAGTCCCCTCTTTGACGCA  984
```

RESULT 6
US-09-790-988-1/c
; Sequence 1, Application US/09790988
; Patent No. 6632935
; GENERAL INFORMATION:
; APPLICANT: SHIGENOBU, SHUJI
; APPLICANT: WATANABE, HIDEMI
; APPLICANT: HATTORI, MASAHIRA
; APPLICANT: SAKAKI, YOSHIYUKI
; TITLE OF INVENTION: GENOME DNA OF BACTERIAL SYMBIONT OF APHIDS
; FILE REFERENCE: 081356/0159
; CURRENT APPLICATION NUMBER: US/09/790,988
; CURRENT FILING DATE: 2001-02-23
; PRIOR APPLICATION NUMBER: JP2000-107160
; PRIOR FILING DATE: 2000-04-07
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 640681
; TYPE: DNA
; ORGANISM: Buchnera sp.
US-09-790-988-1

Query Match   14.6%;   Score 29.4;   DB 4;   Length 640681;
Best Local Similarity 58.6%;   Pred. No. 31;
Matches 51;   Conservative 0;   Mismatches 36;   Indels 0;   Gaps 0;

```
Qy   114  AAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTCCACAAATAAACAA     173
          |||| ||||||| |||| ||
Db 623380 AAAAACAATCTAATTATTATTTTGAGGAAAAAAATGGATTCATACTATTAAAATAAAAA   623321

Qy   174  ACTGTCCGGATTCCTAGAGGGAAGGA     200
          ||||| ||| ||
Db 623320 AAAATCAATTTTCTAAGATGCATGGA   623294
```

RESULT 7
US-08-031-143B-70/c
; Sequence 70, Application US/08031143B
; Patent No. 5518880
; GENERAL INFORMATION:
; APPLICANT: LEONARD, WARREN J.; NOGUCHI, MASAYUKI;
; APPLICANT: MCBRIDE, O. WESLEY
; TITLE OF INVENTION: METHODS FOR DIAGNOSIS AND
; TITLE OF INVENTION: TREATMENT OF XSCID
; NUMBER OF SEQUENCES: 76
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN
; STREET: 345 PARK AVE.
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: USA
; ZIP: 10154
; COMPUTER READABLE FORM:
; MEDIUM TYPE: FLOPPY DISK
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: WORD PERFECT # 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/031,143B
; FILING DATE: 12-MAR-1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:

; NAME: WILLIAM S. FEILER
; REGISTRATION NUMBER: 26,728
; REFERENCE/DOCKET NUMBER: 2026-4061
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; TELEX: 421792
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 513
; TYPE: NUCLEIC ACID
; STRANDEDNESS: SINGLE
; TOPOLOGY: UNKNOWN
; MOLECULE TYPE:
; DESCRIPTION: OLIGONUCLEOTIDE
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: HUMAN
; INDIVIDUAL ISOLATE: IL-2R
US-08-031-143B-70

Query Match   14.5%;   Score 29.2;   DB 1;   Length 513;
Best Local Similarity 62.2%;   Pred. No. 0.6;
Matches 46;   Conservative 0;   Mismatches 28;   Indels 0;   Gaps 0;

```
Qy    12  CCCTGAGCCCTCTTCCTGCCCAACCAGACCAGCAACTTCCAAAATTCTGCCCGTGTTT   71
          ||||| |||||||| |||||||||||||||
Db   477  CCCTAATACCTCTCCTCCCTTCCCATCATGATCCCCTACCTCCTCCTTTCTGCCCATGTAC  418

Qy    72  AGCCTTGTTAAAGG   85
          || |||
Db   417  CCTTATGATAAAAG  404
```

RESULT 8
US-08-916-421B-1
; Sequence 1, Application US/08916421B
; Patent No. 650729
; GENERAL INFORMATION:
; APPLICANT: Bult et al.
; TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococc
; Patent No. 6503729
; TITLE OF INVENTION: jannaschii
; FILE REFERENCE: PB275
; CURRENT APPLICATION NUMBER: US/08/916,421B
; CURRENT FILING DATE: 1997-08-22
; PRIOR APPLICATION NUMBER: US 60/024,428
; PRIOR FILING DATE: 1996-08-22
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1664976
; TYPE: DNA
; ORGANISM: Methanococcus jannaschii
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (28222)..(28222)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (28257)..(28258)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84773)..(84773)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84808)..(84808)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (84812)..(84812)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98120)..(98120)

; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98159)..(98159)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98239)..(98239)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98266)..(98266)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (98343)..(98343)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (103998)..(103998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (148948)..(148948)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (163385)..(163385)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191989)..(191989)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (191995)..(191995)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (231980)..(231980)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234187)..(234187)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234220)..(234220)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (234814)..(234814)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309398)..(309398)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (309418)..(309418)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312837)..(312837)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (312993)..(312993)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (319226)..(319226)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559167)..(559167)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (559241)..(559241)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (600992)..(600992)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (622708)..(622708)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657081)..(657081)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (657203)..(657203)
; OTHER INFORMATION: n equals a, t, c, or g

; NAME/KEY: misc_feature
; LOCATION: (674435)..(674435)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (682442)..(682442)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (713652)..(713652)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (741684)..(741684)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779455)..(779455)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (779676)..(779676)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (855539)..(855539)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (871619)..(871619)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1084830)..(1084830)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1096846)..(1096846)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1119881)..(1119881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1130881)..(1130881)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1310988)..(1310988)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1664854)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g

US-08-916-421B-1

Query Match   14.5%; Score 29.2; DB 4; Length 1664976;
Best Local Similarity   57.8%; Pred. No. 60;
Matches   52; Conservative   0; Mismatches   38; Indels   0; Gaps   0;

Qy   76 TTGTTAAAGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAAGGGGATCTGATTATTGAG 135

Db 793847 TTGATTAATGAGTATGTGTGTATCTCAGTGAAGAGTTAAAATCCTATACAATAAAACAG 793906
Qy 136 GTGTGGAAGAATAAATAATCAGTCCACAA 165
Db 793907 GTTTGTGACAAATATGGGTATAAGTCCAAAA 793936

RESULT 9
US-08-976-259-10
; Sequence 10, Application US/08976259
; Patent No. 616609
GENERAL INFORMATION:
; APPLICANT: Dillon, Patrick J.
; APPLICANT: Choi, Gil H.
; APPLICANT: Welch, Rodney A.
; TITLE OF INVENTION: Nucleotide Sequence of Escherichia coli
Patent No. 616609
; NUMBER OF SEQUENCES: 142
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
; STREET: 1100 New York Ave, N.W., Suite 600
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20005-3934
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/976,259
; FILING DATE: Herewith
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/031,626 AND US 60/061,953
; ATTORNEY/AGENT INFORMATION:
; NAME: Steffe, Eric K.
; REGISTRATION NUMBER: 36,688
; REFERENCE/DOCKET NUMBER: 1488.0740002/EKS/CBM
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 371-2600
; TELEFAX: (202) 371-2540
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2920 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-976-259-10

Query Match 14.2%; Score 28.6; DB 4; Length 2920;
Best Local Similarity 55.6%; Pred. No. 2.8;
Matches 55; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy 99 TTCAGGGAAGTGGGAAAAGGGGGATCTGATTATTGAGGTGTGGAAGGAATAATAATCAG 158
Db 1902 TTTCAGTAATATGGGAGAATAATGAAATGAAATGAATAACACAGACACATTAGAAAAAATAATCAG 1961

Qy 159 TCCACAAATAAACAAACTGTCCCGGGATTCCTAGAGGGAA 197
Db 1962 ACACCAAAAAAACAAAGACCCCGCATATCCTTTCGGGAA 2000

RESULT 10
US-09-221-017B-335
; Sequence 335, Application US/09221017B
; Patent No. 644799
GENERAL INFORMATION:
; APPLICANT: Ross, Bruce C.
; TITLE OF INVENTION: P. GINGIVALIS NUCLEOTIDES AND USES THEREOF
; NUMBER OF SEQUENCES: 1120

CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 755 PAGE MILL ROAD
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304-1018
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FastSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/221,017B
; FILING DATE: 23-DEC-1998
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1182
; FILING DATE: 31-DEC-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP1546
; FILING DATE: 30-JAN-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PP2911
; FILING DATE: 09-APR-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/AU98/01023
; FILING DATE: 10-DEC-1998
; ATTORNEY/AGENT INFORMATION:
; NAME: Monroy, Gladys H
; REGISTRATION NUMBER: 32,430
; REFERENCE/DOCKET NUMBER: 27340-20021.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-813-5600
; TELEFAX: 650-494-0792
; TELEX: 706141
; INFORMATION FOR SEQ ID NO: 335:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3436 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: UNKNOWN
; ORIGINAL SOURCE:
; ORGANISM: PORYPHYROMONAS GINGIVALIS
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 1...3436
US-09-221-017B-335

Query Match 14.0%; Score 28.2; DB 4; Length 3436;
Best Local Similarity 50.4%; Pred. No. 4.3;
Matches 69; Conservative 0; Mismatches 68; Indels 0; Gaps 0;

Qy 50 TTCCAAAATTCTCGCCCGTGTTTAGCCTGTTAAAGGGGGTGTCTCACTCCTTCAGGGAAAG 109
Db 2667 TGCAGAAGTTCTGTTCGACTTTTTAGGTTCCGGAATTTTCTCGTTTTACAGTCAGTCAGCAAA 2726

Qy 110 TGGGAAAAGGGGGATCGATTATTGAGGTGTGGAAGGAATAATAATCAGTCCACAAATAA 169
Db 2727 CGTGGTTCGAGAAAATTTTGTTTGGCGTGGAAAAAATAATTATTATCGAACCAAAACGAA 2786

Qy 170 ACAAACTGTCCGGGATT 186
Db 2787 AAAATTACGCGCGAGT 2803

RESULT 11
US-08-916-421B-1/c
; Sequence 1, Application US/08916421B
; Patent No. 6503729

GENERAL INFORMATION:
APPLICANT: Bult et al.
TITLE OF INVENTION: Complete Genome Sequence of the Methanogenic Archaeon, Methanococc
Patent No. 6503729
FILE REFERENCE: PB275
CURRENT APPLICATION NUMBER: US/08/916,421B
CURRENT FILING DATE: 1997-08-22
PRIOR APPLICATION NUMBER: US 60/024,428
PRIOR FILING DATE: 1996-08-22
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn version 3.1

SEQ ID NO 1
LENGTH: 1664976
TYPE: DNA
ORGANISM: Methanococcus jannaschii
FEATURE:
NAME/KEY: misc_feature
LOCATION: (28222)..(28222)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (28257)..(28258)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84773)..(84773)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84808)..(84808)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (84812)..(84812)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98120)..(98120)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98159)..(98159)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98239)..(98239)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98266)..(98266)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (98343)..(98343)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (103398)..(103398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (148948)..(148948)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (163385)..(163385)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191989)..(191989)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (191995)..(191995)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (231580)..(231980)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234187)..(234187)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234220)..(234220)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (234814)..(234814)

OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309398)..(309398)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (309418)..(309418)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312837)..(312837)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (312993)..(312993)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (319226)..(319226)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559167)..(559167)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (559241)..(559241)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (600992)..(600992)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (622708)..(622708)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657081)..(657081)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (657203)..(657203)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (674435)..(674435)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (682442)..(682442)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (713652)..(713652)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (741684)..(741684)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779455)..(779455)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (779676)..(779676)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (855539)..(855539)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (871619)..(871619)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1084830)..(1084830)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1096846)..(1096846)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1119881)..(1119881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1130881)..(1130881)
OTHER INFORMATION: n equals a, t, c, or g
NAME/KEY: misc_feature
LOCATION: (1310988)..(1310988)
OTHER INFORMATION: n equals a, t, c, or g

; NAME/KEY: misc_feature
; LOCATION: (1313224)..(1313224)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349473)..(1349473)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1349491)..(1349491)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1470091)..(1470091)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1569020)..(1569020)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1602912)..(1602912)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1603734)..(1603734)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1637998)..(1637998)
; OTHER INFORMATION: n equals a, t, c, or g
; NAME/KEY: misc_feature
; LOCATION: (1664855)..(1664855)
; OTHER INFORMATION: n equals a, t, c, or g
US-08-916-421B-1

Query Match   14.0%; Score 28.2; DB 4; Length 1664976;
Best Local Similarity   55.7%; Pred. No. 1.2e+02;
Matches   54; Conservative   0; Mismatches   43; Indels   0; Gaps   0;

Qy   14 CTGAAGCCCTCTTCCCTGCCCAACCAGACCAACTTCCAAAATTCTGCCCGTGTTTAG 73
            |||| |||||||||||| ||||| |||||||||||| |||||||| |||||||||
Db   1193490 CTGATAATGTCTTTCCACTCCCAGTAGGTGAGCAAAATTAAAACATTCTTCCCTTCATGAA 1193431

Qy   74 CCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGT 110
            | |||| | ||| | || ||| | || ||| |||
Db   1193430 TCTCTTTAATTGCATATCTTTGTGTGGAGTAAAGGT 1193394

RESULT 12
US-09-439-313-239
; Sequence 239, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqui
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 239
; LENGTH: 239
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-439-313-239

Query Match   13.9%; Score 28; DB 4; Length 239;
Best Local Similarity   48.7%; Pred. No. 1.1;

Matches   76; Conservative   0; Mismatches   80; Indels   0; Gaps   0;

Qy   41 ACCAGCAACTTCCAAAATTCTGCCCGTGTTTAGCCTTGTTAAAGGGGTGTCTCACTCCTT 100
            ||| ||| ||| |||||||||||||||||||||||| ||||||||||||||||||||||||
Db   38 AACATAAAGTTCACATAACTGCTTCGTCGTCAAACCATGATACTGAGCTTTGTGACAACCC 97

Qy   101 CAGGGAAAGTGGGAAAAGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAAATCAGTC 160
            ||| |||| ||||||| || | | ||| | ||||||||||| | | || |||||||
Db   98 AGAAATAACTAAGAGAGAGGCAAACATAATACCTTAGAGATCAAGAAACATTTACACAGTT 157

Qy   161 CACAAATAAACAAACTGTCCGGGATTCCTAGAGGGA 196
            ||||||| ||| ||| | || ||| ||||| | |
Db   158 CAACTGTTTAAAAATAGCTCAACATTCAGCCAGTGA 193

RESULT 13
US-09-352-616A-239
; Sequence 239, Application US/09352616A
; Patent No. 6395278
; GENERAL INFORMATION:
; APPLICANT: Dillon, Davin C.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang, Yuqui
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; TITLE OF INVENTION: OF PROSTATE CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.427C8
; CURRENT APPLICATION NUMBER: US/09/352,616A
; CURRENT FILING DATE: 1999-07-13
; NUMBER OF SEQ ID NOS: 472
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 239
; LENGTH: 239
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-352-616A-239

Query Match   13.9%; Score 28; DB 4; Length 239;
Best Local Similarity   48.7%; Pred. No. 1.1;
Matches   76; Conservative   0; Mismatches   80; Indels   0; Gaps   0;

Qy   41 ACCAGCAACTTCCAAAATTCTGCCCGTGTTTAGCCTTGTTAAAGGGGTGTCTCACTCCTT 100
            ||| ||| ||| |||||||||||||||||||||||| ||||||||||||||||||||||||
Db   38 AACATAAAGTTCACATAACTGCTTCGTCGTCAAACCATGATACTGAGCTTTGTGACAACCC 97

Qy   101 CAGGGAAAGTGGGAAAAGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAAATCAGTC 160
            ||| |||| ||||||| || | | ||| | ||||||||||| | | || |||||||
Db   98 AGAAATAACTAAGAGAGAGGCAAACATAATACCTTAGAGATCAAGAAACATTTACACAGTT 157

Qy   161 CACAAATAAACAAACTGTCCGGGATTCCTAGAGGGA 196
            ||||||| ||| ||| | || ||| ||||| | |
Db   158 CAACTGTTTAAAAATAGCTCAACATTCAGCCAGTGA 193

RESULT 14
US-09-232-149A-239
; Sequence 239, Application US/09232149A
; Patent No. 6465611
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
; TITLE OF INVENTION: CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.427C6
; CURRENT APPLICATION NUMBER: US/09/232,149A
; CURRENT FILING DATE: 1999-01-15
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 239
; LENGTH: 239
; TYPE: DNA

; ORGANISM: Homo sapien
US-09-232-149A-239

Query Match   13.9%;   Score 28;   DB 4;   Length 239;
Best Local Similarity   48.7%;   Pred. No. 1.1;
Matches   76;   Conservative   0;   Mismatches   80;   Indels   0;   Gaps   0;

Qy   41 ACCAGCAACTTCCAAAATTCTGCCGTGTTTAGCCTTGTTAAAGGGGTGTCTCACTCCTT 100
            ||||||||| ||| |||| ||||||| ||| ||| |||||| ||||||||||
Db   38 AACATAAAAGTTCACATAACTGCTTCTGTCAAACCATGATACTGAGCTTTGTGACAACCC 97

Qy   101 CAGGGAAAGTGGGAAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTC 160
                       ||| |||
Db   98 AGAAATAACTAAGAGAAGGCAAACATAATACCTTAGAGATCAAGAAACATTTACACAGTT 157

Qy   161 CACAAATAAACAAACTGTCCGGGATTCCTAGAGGGA 196
            ||| ||||| |||| |||
Db   158 CAACTGTTTAAAAATAGCTCAACATTCAGCCAGTGA 193

RESULT 15
US-09-159-812-239
; Sequence 239, Application US/09159812A
; Patent No. 6613872
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF
; TITLE OF INVENTION: PROSTATE CANCER AND METHODS FOR THEIR USE
; FILE REFERENCE: 210121.428C5
; CURRENT APPLICATION NUMBER: US/09/159,812A
; CURRENT FILING DATE: 1998-09-23
; NUMBER OF SEQ ID NOS: 306
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 239
;   LENGTH: 239
;   TYPE: DNA
;   ORGANISM: Homo sapien
US-09-159-812-239

Query Match   13.9%;   Score 28;   DB 4;   Length 239;
Best Local Similarity   48.7%;   Pred. No. 1.1;
Matches   76;   Conservative   0;   Mismatches   80;   Indels   0;   Gaps   0;

Qy   41 ACCAGCAACTTCCAAAATTCTGCCGTGTTTAGCCTTGTTAAAGGGGTGTCTCACTCCTT 100
            ||||||||| ||| |||| ||||||| ||| ||| |||||| ||||||||||
Db   38 AACATAAAAGTTCACATAACTGCTTCTGTCAAACCATGATACTGAGCTTTGTGACAACCC 97

Qy   101 CAGGGAAAGTGGGAAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTC 160
                       ||| |||
Db   98 AGAAATAACTAAGAGAAGGCAAACATAATACCTTAGAGATCAAGAAACATTTACACAGTT 157

Qy   161 CACAAATAAACAAACTGTCCGGGATTCCTAGAGGGA 196
            ||| ||||| |||| |||
Db   158 CAACTGTTTAAAAATAGCTCAACATTCAGCCAGTGA 193

Search completed: March 25, 2004, 15:34:33
Job time : 43.5358 secs

OM nucleic - nucleic search, using sw model

Copyright GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on: March 25, 2004, 10:03:04 ; Search time 174.116 Seconds
(without alignments)
4297.861 Million cell updates/sec

Title: US-09-963-285-1_COPY_223_423
Perfect score: 201
Sequence: 1 acaaatgttctccctgaagc........ggattccctgagaggaggaggag 201

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 258946 seqs, 1861504846 residues
Total number of hits satisfying chosen parameters: 4917892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10C_PUBCOMB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Match	Query Length	DB	ID	Description
	1	201	100.0	6458	9	US-09-963-285-1	Sequence 1, Appli
	2	66	32.8	6021	9	US-09-963-285-5	Sequence 5, Appli
	3	35.6	17.7	477	12	US-10-424-599-39980	Sequence 39980, A
c	4	33	16.4	263744	14	US-10-229-834A-6	Sequence 6, Appli
c	5	31	15.4	1606	12	US-10-424-599-93300	Sequence 93300, A
	6	31	15.4	1830	12	US-10-302-172-616	Sequence 616, App
c	7	30.8	15.3	360	9	US-09-960-352-6958	Sequence 6958, Ap
	8	30.8	15.3	2443	12	US-10-425-114-12307	Sequence 32307, A
	9	30.4	15.1	491	9	US-09-864-761-14243	Sequence 14243, A
c	10	30.4	15.1	3574	9	US-09-912-917-1	Sequence 1, Appli
c	11	29.6	14.7	142299	10	US-09-911-077A-14	Sequence 14, Appl
	12	29.4	14.6	624	15	US-10-027-632-194870	Sequence 194870,
	13	29.4	14.6	1680	15	US-10-027-632-98041	Sequence 98041, A
	14	29.4	14.6	1680	15	US-10-027-632-98042	Sequence 98042, A
	15	29.4	14.6	1680	15	US-10-027-632-98043	Sequence 98043, A
c	16	29.4	14.6	2103	15	US-10-104-047-1788	Sequence 1788, Ap
c	17	29.4	14.6	2592	15	US-10-108-260A-444	Sequence 444, App
c	18	29.4	14.6	68001	16	US-10-444-206-297	Sequence 297, App
c	19	29.2	14.6	640681	9	US-09-790-988-1	Sequence 1, Appli
c	20	29.2	14.5	394	12	US-10-424-599-67913	Sequence 67913, A
	21	29	14.4	6943	15	US-10-292-798-1229	Sequence 1229, Ap
	22	29	14.4	7058	16	US-10-257-166-133	Sequence 133, App
c	23	29	14.4	177556	10	US-09-952-213D-6	Sequence 6, Appli
c	24	29	14.4	822900	15	US-10-292-798-1393	Sequence 1393, Ap
	25	29	14.4	3673778	14	US-10-312-841-2	Sequence 2, Appli
	26	28.8	14.3	541	15	US-10-027-632-321278	Sequence 321278,
c	27	28.8	14.3	791	12	US-10-424-599-15773	Sequence 15773, A
	28	28.8	14.3	797	15	US-10-027-632-157250	Sequence 157250,
c	29	28.8	14.3	1531	14	US-10-106-698-2012	Sequence 2012, Ap
	30	28.8	14.3	2066	10	US-09-814-353-19759	Sequence 19759, A
c	31	28.8	14.3	2205	15	US-10-104-047-79	Sequence 79, Appl
c	32	28.8	14.3	2761	12	US-10-424-599-31015	Sequence 31015, A
	33	28.8	14.3	5503	13	US-10-044-090-7	Sequence 7, Appli
	34	28.8	14.3	19653	14	US-10-311-455-1308	Sequence 1308, Ap
	35	28.8	14.3	1691139	14	US-10-067-514-1	Sequence 1, Appli
	36	28.8	14.3	1691139	15	US-10-419-723-1	Sequence 1, Appli
c	37	28.6	14.2	441	9	US-09-770-444-905	Sequence 905, App
	38	28.6	14.2	2920	9	US-09-956-004-10	Sequence 10, Appl
	39	28.6	14.2	6115	9	US-09-292-758-131	Sequence 131, App
	40	28.6	14.2	45980	10	US-09-957-956-6	Sequence 6, Appli
	41	28.6	14.2	358846	15	US-10-292-798-1095	Sequence 1095, Ap
c	42	28.4	14.1	261	15	US-10-085-783A-3436	Sequence 3436, Ap
c	43	28.4	14.1	261	15	US-10-242-535A-3436	Sequence 3436, Ap
	44	28.4	14.1	1143	14	US-10-032-585-6514	Sequence 6514, Ap
c	45	28.4	14.1	17239	9	US-09-764-869-2326	Sequence 2326, Ap

ALIGNMENTS

RESULT 1
US-09-963-285-1
; Sequence 1, Application US/09963285
; Patent No. US20020090707A1
; GENERAL INFORMATION:
; APPLICANT: Enerbck, Sven
; APPLICANT: Krook, Katarina
; APPLICANT: Rondahl, Lena
; APPLICANT: Wasserman, Wyeth
; TITLE OF INVENTION: PROMOTER SEQUENCES
; FILE REFERENCE: 13425-042001
; CURRENT APPLICATION NUMBER: US/09/963,285
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: SE 0004102-0
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/238,897
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: SE 0003435-5
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 6458
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (2235)...(3737)
US-09-963-285-1

Query Match 100.0%; Score 201; DB 9; Length 6458;
Best Local Similarity 100.0%; Pred. No. 6.8e-57;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy    1 ACAAATGTTCTCCCTGAAGCCCTCTTCCCTGCCCAACCAGACCAACTTCCAAAATTC  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  223 ACAAATGTTCTCCCTGAAGCCCTCTTCCCTGCCCAACCAGACCAACTTCCAAAATTC 282
```

```
Qy   61 TGCCGGTGTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAAGGG 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  283 TGCCGGTGTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAAGGG 342

Qy  121 GATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTCCACAAATAAACAAACTGTCC 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  343 GATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTCCACAAATAAACAAACTGTCC 402

Qy  181 GGGATTCCTAGAGGGAAGGAG 201
        |||||||||||||||||||||
Db  403 GGGATTCCTAGAGGGAAGGAG 423
```

RESULT 2
US-09-963-285-5
; Sequence 5, Application US/09963285
; Patent No. US2002009070A1
; GENERAL INFORMATION:
; APPLICANT: Enerbck, Sven
; APPLICANT: Krook, Katarina
; APPLICANT: Rondahl, Lena
; APPLICANT: Wasserman, Wyeth
; TITLE OF INVENTION: PROMOTER SEQUENCES
; FILE REFERENCE: 13425-042001
; CURRENT APPLICATION NUMBER: US/09/963,285
; CURRENT FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: SE 0004102-0
; PRIOR FILING DATE: 2000-11-09
; PRIOR APPLICATION NUMBER: US 60/238,897
; PRIOR FILING DATE: 2000-10-10
; PRIOR APPLICATION NUMBER: SE 0003435-5
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 6021
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: exon
; LOCATION: (1649)...(438)
US-09-963-285-5

Query Match   32.8%;   Score 66; DB 9; Length 6021;
Best Local Similarity 73.1%;   Pred. No. 2.2e-11;
Matches 155;   Conservative 0;   Mismatches 45;   Indels 12;   Gaps 5;

```
Qy    1 ACAAATGTCTCCCTGAAGCCCTTCCCTGCCCAACCAGACCAACTTCCAAAATTC 60
        |||||||   |   |||| ||  |||||  | ||  ||||| |||| |||||||
Db  195 ACAAATGTTTTATCTGTCGCCTTCTTCCCTACCCACCGGACCAACAACTTCCAGAGGT 254

Qy   61 TGCCCGTG---TTTAGCCTGTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAA 117
        |||| |||   |   |||  |  ||   |||||  ||  |||  |||||||||| ||  |
Db  255 TCTGCGAGGCATAGAGCCATTCCGTAGGGACATCTCGGTGCTTCTGAGGAAGCGGACCGA 314

Qy  118 G--GGGATCTGATTA-----TTGAGGTGTGGAAGGAATAAATAATCAGTCCACAAATAAAC 171
        |  |||||| ||||      ||||| |||||||||||||||| |||||||||||||||||
Db  315 GCAGGGATCCGATGACGACCTGGAGATGTTGAAGGAATAAAT-ACCAGTCCACAAATAAAC 373

Qy  172 AAACTGT--CCGGGATTCCTAGAGGGAAGGAG 201
        |||||||   ||||||||||||||||||||||
Db  374 AAACTGTCCCCGGGATTCCTAGAGGGAAGGAG 405
```

RESULT 3
US-10-424-599-39980/c
; Sequence 39980, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 39980
; LENGTH: 477
; TYPE: DNA
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_136101C.1
US-10-424-599-39980

Query Match   17.7%;   Score 35.6; DB 12; Length 477;
Best Local Similarity 51.2%;   Pred. No. 0.13;   Indels 0;   Gaps 0;
Matches 83;   Conservative 0;   Mismatches 79;

```
Qy   36 ACCAGACCAGCAACTTCCAAAATTCTGCCCGTGTTAGCCTTGTTAAAGGGGTGTCTCAC 95
        || |  ||| |||   ||||| ||||  ||| || | | |||  |||   |  ||| |
Db  383 ACAACAACATGAATAACAAAAGGAATGAGATTTTCATTCTTCGAAAATTAATGACTTAA 324

Qy   96 TCCTTCAGGGAAAGTGGGAAAAGGGGATCCGATCTGATTATTGAGGTGTGGAAGGAATAAAT 155
        | |||||  ||| | |  |||| | |||| |  ||| ||| | |   |  |    ||||| |
Db  323 ACATTCAATGAAATCAGGATGCAGGGATCTTATTATAAAATAATTGACACATTAAAAT 264

Qy  156 CAGTCCACAAATAAACAAACTGTCCGGGATTCCTAGAGGGAA 197
        | || |||||| |  |||| | |    |||  | |||| |||
Db  263 TGAAAAACCAAAATTGTACAGCTACGGTATCCAAAGACTGAA 222
```

RESULT 4
US-10-229-834A-6/c
; Sequence 6, Application US/10229834A
; Publication No. US2003015003A1
; GENERAL INFORMATION:
; APPLICANT: Lawrence Berkeley National Laboratory
; APPLICANT: Rubin, Edward
; APPLICANT: Pennacchio, Len
; TITLE OF INVENTION: A novel apolipoprotein gene involved in lipid metabolism
; FILE REFERENCE: IB-1709
; CURRENT APPLICATION NUMBER: US/10/229,834A
; CURRENT FILING DATE: 2002-08-27
; PRIOR APPLICATION NUMBER: US 60/318,219
; PRIOR FILING DATE: 2001-08-27
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 263744
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (750)..(850)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1584)..(1683)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2154)..(2154)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2358)..(2457)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3208)..(3307)
; OTHER INFORMATION: gap of unknown length
; FEATURE:
; NAME/KEY: misc_feature

```
      LOCATION: (4044)..(4143)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (4809)..(4908)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (4921)..(4921)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (5652)..(5751)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (6469)..(6469)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (6477)..(6477)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (6482)..(6482)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (6487)..(6487)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (6489)..(6490)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (6523)..(6622)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (6637)..(6637)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (7410)..(7509)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (8177)..(8177)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (8248)..(8347)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (8358)..(8358)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (8967)..(8967)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (9057)..(9057)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (9105)..(9204)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (9980)..(10079)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (10092)..(10092)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (10827)..(10926)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (11701)..(11800)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (11816)..(11816)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (11827)..(11827)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (11829)..(11829)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (12565)..(12664)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (13327)..(13327)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (13440)..(13539)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (13554)..(13554)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (14133)..(14133)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (14213)..(14213)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (14293)..(14392)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (14559)..(14559)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (14900)..(14900)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (15119)..(15218)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (15231)..(15233)
      OTHER INFORMATION: gap of unknown length
  FEATURE:
      NAME/KEY: misc_feature
      LOCATION: (15919)..(15919)
      OTHER INFORMATION: gap of unknown length
```

```
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15922)..(15922)
;   OTHER INFORMATION: gap of unknown length
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (15962)..(16061)
;   OTHER INFORMATION: gap of unknown length
;   FEATURE:
;   NAME/KEY: misc_feature
;   LOCATION: (16757)..(16757)
;   OTHER INFORMATION: gap of unknown length
;   FEATURE:
;   NAME/KEY: misc_feature

  Query Match        16.4%;   Score 33;   DB 14;   Length 263744;
  Best Local Similarity  65.8%;   Pred. No. 16;
  Matches   48;   Conservative   0;   Mismatches   25;   Indels   0;   Gaps   0;

Qy      102  AGGGAAAGTGGGAAAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTCC  161
             ||||||||| ||||||         ||||||||||||| |||||||||     |||||||
Db    52915  AGGGAAAATCAGGAAAAGTGTTTAAAATGTTTGATGTAGGAGGAAATGAATAAATAAGTTA  52856

Qy      162  ACAAATAAACAAA  174
             ||||| ||||||
Db    52855  ATAGGAAAAAAAA  52843

RESULT 5
US-10-424-599-93300
; Sequence 93300, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
;   APPLICANT: La Rosa Thomas J
;   APPLICANT: Kovalic David K
;   APPLICANT: Zhou Yihua
;   APPLICANT: Cao Yongwei
;   TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
;   TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
;   FILE REFERENCE: 38-21(53223)B
;   CURRENT APPLICATION NUMBER: US/10/424,599
;   CURRENT FILING DATE: 2003-04-28
;   NUMBER OF SEQ ID NOS: 285684
;   SEQ ID NO 93300
;   LENGTH: 606
;   TYPE: DNA
;   ORGANISM: Glycine max
;   FEATURE:
;   OTHER INFORMATION: Clone ID: PAT_MRT3847_55264C.1
US-10-424-599-93300

  Query Match        15.4%;   Score 31;   DB 12;   Length 606;
  Best Local Similarity  57.9%;   Pred. No. 5.1;
  Matches   55;   Conservative   0;   Mismatches   40;   Indels   0;   Gaps   0;

Qy      103  GGGAAAGTGGGAAAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTCCA  162
             |||||| ||||||||||||||      |||||||||||| ||| |||||| |||||||||
Db      486  GGGAAAGGGGGAAAAAGGGGGGGTTTAACCGGGCGTGGTTTTAAGAGAAGAATAGAGAAAA  545

Qy      163  CAAATAAACAAACTGTCCGGGATTCCTAGAGGGAA  197
             ||||| |||||                ||||||||
Db      546  GAGGTACACAAAGATTTGGGGCCCCCAACAGGGAA  580

RESULT 6
US-10-302-172-616
; Sequence 616, Application US/10302172
; Publication No. US20040053250A1
; GENERAL INFORMATION:
;   APPLICANT: Tang, Y. Tom
;   APPLICANT: Xue, Aidong J.
;   APPLICANT: Drmanac, Radoje T.
;   TITLE OF INVENTION: No. US20040053250A1e1 Arginine-rich Protein-like Nucleic Acids an
```

```
;   TITLE OF INVENTION:  Polypeptides
;   FILE REFERENCE: 803_1CNCP
;   CURRENT APPLICATION NUMBER: US/10/302,172
;   CURRENT FILING DATE: 2002-11-21
;   PRIOR APPLICATION NUMBER: US 10/225,251
;   PRIOR FILING DATE: 2002-08-20
;   PRIOR APPLICATION NUMBER: PCT US02/05095
;   PRIOR FILING DATE: 2002-03-05
;   PRIOR APPLICATION NUMBER: US 09/799,451
;   PRIOR FILING DATE: 2001-03-05
;   NUMBER OF SEQ ID NOS: 950
;   SOFTWARE: pt_FL_genes Version 2.0
;   SEQ ID NO 616
;   LENGTH: 1830
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (337)..(696)
US-10-302-172-616

  Query Match        15.4%;   Score 31;   DB 12;   Length 1830;
  Best Local Similarity  56.3%;   Pred. No. 8.3;
  Matches   58;   Conservative   0;   Mismatches   45;   Indels   0;   Gaps   0;

Qy       76  TTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAAGGGGATCTGATTATTGAG  135
             ||||||||        |||| ||||||||||||||||||||||||||||| ||||||||
Db      640  TTTTTTAAGGGGGTTCAGATCCTGGAAGAAAAATGACAATAGTGACTATGAGAATGTAG  699

Qy      136  GTGTGGAAGGAATAAATAATCAGTCCACAAATAAACAAACTGT  178
             |||||||||||||||||||| ||||||||||||||| ||||||
Db      700  AAGTGGCATCTGAAAACAATTGGTACCAGTAAACTGGCAGT    742

RESULT 7
US-09-960-352-6958/c
; Sequence 6958, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
;   APPLICANT: Warren, Wesley C.
;   APPLICANT: Tao, Nengbing
;   APPLICANT: Byatt, John C.
;   APPLICANT: Mathialagan, Nagappan
;   TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
;   TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
;   FILE REFERENCE: 16511.006/37-21(10298)C
;   CURRENT APPLICATION NUMBER: US/09/960,352
;   CURRENT FILING DATE: 2001-09-24
;   NUMBER OF SEQ ID NOS: 15112
;   SEQ ID NO 6958
;   LENGTH: 360
;   TYPE: DNA
;   ORGANISM: Bos taurus
;   OTHER INFORMATION: Clone ID: 30-LIB3057-020-Q1-K1-H5
US-09-960-352-6958

  Query Match        15.3%;   Score 30.8;   DB 9;   Length 360;
  Best Local Similarity  55.7%;   Pred. No. 4.8;
  Matches   59;   Conservative   0;   Mismatches   47;   Indels   0;   Gaps   0;

Qy       54  AAAATTCTGCCCGTGTTTAGCCTTGTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGG  113
             ||||||       |||| |       |||||| |||||||||||||||||||||||||||
Db      347  AAAATTTTATTCTCATTAAACTTTTTAAATGGGTATAAAATTTTGTGACAGATTTTTGG  288

Qy      114  AAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAATAATCAGT  159
             |||||||||||                |||||||||||||||||
Db      287  AAAAGTTGTTTTCATTAAAAAGTACTGATTTTAAAAACTAATAATT  242

RESULT 8
US-10-425-114-32307
; Sequence 32307, Application US/10425114
; Publication No. US20040034888A1
```

GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated with
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
SEQ ID NO 32307
; LENGTH: 2443
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: UC-ZMFLB73291G02_FLI
US-10-425-114-32307

Query Match　　　　15.3%;　　Score 30.8;　　DB 12;　　Length 2443;
Best Local Similarity　55.7%;　　Pred. No. 11;
Matches　59;　Conservative　0;　Mismatches　47;　Indels　0;　Gaps　0;

Qy　　91 CTCCACTCCTTCAGGGAAAGTGGGAAAAGGGGATCTGATTATTGAGGTGTGGAGGAATAA 150
　　　　　||||| ||| || ||| |||| |||| ||| || || | |||||| ||| |||| |||
Db　1425 CTTACATACACCGGTCAGGTCGGACAGGGAGGCAGGTAAATACTGGTGTTGCTGTCATGC 1484

Qy　　151 ATAATCAGTCCACAAATAAACAAACTGTCCGGGATTCCTAGAGGGGA 196
　　　　　　|| ||||　　||||||||||　|||||
Db　1485 TTTATGAGCCCAGATATTAAGCACAGTGTCAGCAGACTAGAAAGGGA 1530

RESULT 9
US-09-864-761-14243
; Sequence 14243, Application US/09864761
; Patent No. US2002004876361
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 14243
; LENGTH: 491
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC009893.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.92
US-09-864-761-14243

Query Match　　　　15.1%;　　Score 30.4;　　DB 9;　　Length 491;
Best Local Similarity　71.4%;　　Pred. No. 7.5;
Matches　40;　Conservative　0;　Mismatches　16;　Indels　0;　Gaps　0;

Qy　　119 GGGATCTGATTATTGAGGTGTGGAGGAATAAATAATCAGTCCACAAATAAACAAA 174
　　　　　 ||| |||||||||||||||||||| || ||||||||| ||||| || ||||||||
Db　　　1 GGGATCTGATTAATGAGGGATTGTAGGAGTAAGAAGAACTCTAAAATATAGAAA 56

RESULT 10
US-09-912-917-1
; Sequence 1, Application US/09912917
; Patent No. US2002028487A1
; GENERAL INFORMATION:
; APPLICANT: LaThangue, Nicholas B
; APPLICANT: Shikama, No. US2002002028487A1ko
; APPLICANT: The University Court of the University of Glasgow
; TITLE OF INVENTION: JMY, a co-activator for p300/CBP, nucleic acid encoding
; TITLE OF INVENTION: JMY and uses thereof
; FILE REFERENCE: AHB/LP5731500
; CURRENT APPLICATION NUMBER: US/09/912,917
; CURRENT FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 09/311,236
; PRIOR FILING DATE: 1999-05-13
; PRIOR APPLICATION NUMBER: GB 9818235.5
; PRIOR FILING DATE: 1998-08-20
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
; LENGTH: 3574
; TYPE: DNA
; ORGANISM: Mus sp.
US-09-912-917-1

Query Match　　　　15.1%;　　Score 30.4;　　DB 9;　　Length 3574;
Best Local Similarity　52.3%;　　Pred. No. 18;
Matches　67;　Conservative　0;　Mismatches　61;　Indels　0;　Gaps　0;

Qy　　9 TCTCCCTGAAGCCCTCTTCCCTGCCAACCAGCAGACCAACTTCAAAATTCCAAAATTCTGCCCCGTG 68
　　　　| |||||||||||||||||||||| | || ||| |||| |||| ||||||| || || |||||
Db　2520 TCTTCCTCCAACACCTCCCCTCCCCCCCCCACCCCCCACCTCCACCCCCACCCCCTGCC 2579

Qy　　69 TTTAGCCTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAAGGGGATCTGAT 128
　　　　| |||| || ||| || ||| ||| ||||| |||||||
Db　2580 TGTTGCAAAGGACAATGGGGCCTCCACCACTGCAGACACTGGAGAAGATGCACTTAG 2639

Qy　　129 TATTGAGG 136
　　　　　 |||||
Db　2640 GACGGAGG 2647

RESULT 11
US-09-911-077A-14/c
; Sequence 14, Application US/09911077A
; Publication No. US20030114399A1
; GENERAL INFORMATION:
; APPLICANT: BLAKELY, RANDY D.
; APPLICANT: APPARSUNDARAM, SUBRAMANIAM
; APPLICANT: FERGUSON, SHAWN
; TITLE OF INVENTION: HUMAN AND MOUSE CHOLINE TRANSPORTER cDNA
; FILE REFERENCE: VBLT:008US
; CURRENT APPLICATION NUMBER: US/09/911,077A
; CURRENT FILING DATE: 2001-07-23
; NUMBER OF SEQ ID NOS: 27
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 142299
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Primer
; FEATURE:
; NAME/KEY: modified base
; LOCATION: (1305)..(127835)
; OTHER INFORMATION: N = A, C, G or T/U
US-09-911-077A-14

Query Match   14.7%; Score 29.6; DB 10; Length 142299;
Best Local Similarity   57.6%; Pred. No. 1.7e+02;
Matches   53; Conservative   0; Mismatches   39; Indels   0; Gaps   0;

Qy   11 TCCCTGAGCCCTCTTCCCTGCCCAACCAGACCAGCAACTTCCAAAATTCTGCCCGTGTT 70
          ||| |||||||||||| ||||||||| ||| |||| |||||| |||| |
Db 42105 TCCTGGGTGCCTACACCATTTCCTAATGAAACTTGAAGATTTGACCATTCGTCCATTTC 42046

Qy   71 TAGCCTTGTTAAAGGGGTGTCTCACTCCTTCA 102
          |||||||||| ||| ||||||| ||||
Db 42045 CATCCTTGTTTGACGTGGTCTATAAGTGTTTTA 42014
US-09-911-077A-14

RESULT 12
US-10-027-632-194870
; Sequence 194870, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 194870
; LENGTH: 624
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-194870

Query Match   14.6%; Score 29.4; DB 15; Length 624;
Best Local Similarity   54.1%; Pred. No. 18;
Matches   60; Conservative   0; Mismatches   51; Indels   0; Gaps   0;

Qy   39 AGACCAGCAACTTCCAAATTCTGCCCGTGTTAGCCTTGTTAAAGGGGTGTCTCACTCC 98
          || ||||| ||||||||||| ||||||||| || |||| ||||||||||| |||
Db   28 AGAAGAGTTATTTCACACCTTCTTCCTGTGCTTTCCCAGGGTAACGGGGCTGTCTCATTTA 87

Qy   99 TTCAGGGAAAGTGGGAAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATA 149
          || | |||| ||||||| ||| |||||| |||
Db   88 CTGAGTACCACCACAGGGGTGTAACTCTGAATTGGGGACTTGGGAAATGTAGA 138

RESULT 13
US-10-027-632-98041
; Sequence 98041, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 98041
; LENGTH: 1680
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-98041

Query Match   14.6%; Score 29.4; DB 15; Length 1680;
Best Local Similarity   60.8%; Pred. No. 28;
Matches   48; Conservative   0; Mismatches   31; Indels   0; Gaps   0;

Qy   114 AAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAAATAATCAGTCCACAAATAAACAA 173
          |||||| ||||||||| || |||||| | || || |||| ||
Db   1169 AAAAGAAGATCTGGTTCTTATGGAGAGCCTAGAGTGAGTACAGCAGCAGTGCAGTGAAAAATAT 1228

Qy   174 ACTGTCCGGGATTCCTAGA 192
          ||| ||||| ||||
Db   1229 TTAATCCTGAATACTTAGA 1247

RESULT 14
US-10-027-632-98042
; Sequence 98042, Application US/10027632
; Publication No. US20030204075A9
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676

```
;   PRIOR FILING DATE: 2000-04-20
;   PRIOR APPLICATION NUMBER: US 60/193,483
;   PRIOR FILING DATE: 2000-03-29
;   PRIOR APPLICATION NUMBER: US 60/185,218
;   PRIOR FILING DATE: 2000-02-24
;   PRIOR APPLICATION NUMBER: US 60/167,363
;   PRIOR FILING DATE: 1999-11-23
;   PRIOR APPLICATION NUMBER: US 60/156,358
;   PRIOR FILING DATE: 1999-09-28
;   PRIOR APPLICATION NUMBER: US 60/146,002
;   PRIOR FILING DATE: 1999-08-09
;   NUMBER OF SEQ ID NOS: 325720
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 98042
;     LENGTH: 1680
;     TYPE: DNA
;     ORGANISM: Human
US-10-027-632-98042

Query Match             14.6%;  Score 29.4;  DB 15;  Length 1680;
Best Local Similarity   60.8%;  Pred. No. 28;
Matches   48;  Conservative  0;  Mismatches  31;  Indels  0;  Gaps  0;

Qy        114 AAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTCCACAAATAAACAA 173
              ||||   |||||||| || |||     || || |  |  || |||||||||||| || |||
Db       1169 AAAAGAAGATCTCGTTCTTATGGAGAGGCTAGATACAGCAGCAGTGCAGTGAAAAATAT 1228

Qy        174 ACTGTCCGGGATTCCTAGA 192
              |||| ||||||||||||||| || ||||
Db       1229 TTAATCCTGAATACTTAGA 1247


RESULT 15
US-10-027-632-98043
;   Sequence 98043, Application US/10027632
;   Publication No. US20030204075A9
;   GENERAL INFORMATION:
;   APPLICANT: Wang, David G.
;   TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
;   TITLE OF INVENTION:  Polymorphisms in the Human Genome
;   FILE REFERENCE: 108827.129
;   CURRENT APPLICATION NUMBER: US/10/027,632
;   CURRENT FILING DATE:  2002-04-30
;   PRIOR APPLICATION NUMBER: US 60/218,006
;   PRIOR FILING DATE: 2000-07-12
;   PRIOR APPLICATION NUMBER: US 60/198,676
;   PRIOR FILING DATE: 2000-04-20
;   PRIOR APPLICATION NUMBER: US 60/193,483
;   PRIOR FILING DATE: 2000-03-29
;   PRIOR APPLICATION NUMBER: US 60/185,218
;   PRIOR FILING DATE: 2000-02-24
;   PRIOR APPLICATION NUMBER: US 60/167,363
;   PRIOR FILING DATE: 1999-11-23
;   PRIOR APPLICATION NUMBER: US 60/156,358
;   PRIOR FILING DATE: 1999-09-28
;   PRIOR APPLICATION NUMBER: US 60/146,002
;   PRIOR FILING DATE: 1999-08-09
;   NUMBER OF SEQ ID NOS: 325720
;   SOFTWARE: FastSEQ for Windows Version 4.0
;   SEQ ID NO 98043
;     LENGTH: 1680
;     TYPE: DNA
;     ORGANISM: Human
US-10-027-632-98043

Query Match             14.6%;  Score 29.4;  DB 15;  Length 1680;
Best Local Similarity   60.8%;  Pred. No. 28;
Matches   48;  Conservative  0;  Mismatches  31;  Indels  0;  Gaps  0;

Qy        114 AAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTCCACAAATAAACAA 173
              ||||   |||||||| || |||     || || |  |  || |||||||||||| || |||
Db       1169 AAAAGAAGATCTCGTTCTTATGGAGAGGCTAGATACAGCAGCAGTGCAGTGAAAAATAT 1228

Qy        174 ACTGTCCGGGATTCCTAGA 192
              |||| ||||||||||||||| || ||||
Db       1229 TTAATCCTGAATACTTAGA 1247


Search completed: March 25, 2004, 15:52:59
Job time : 178.116 secs
```

OM nucleic - nucleic search, using sw model

Copyright　GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

Run on:　　March 25, 2004, 09:53:09 ; Search time 1518.97 Seconds
　　　　　　　　　　　　　　　　　　　　(without alignments)
　　　　　　　　　　　　　　　　　　　　3951.570 Million cell updates/sec

Title:　　　US-09-963-285-1_COPY_223_423
Perfect score:　201
Sequence:　1 acaaatgttctccctgaagc.........ggattcctagaggggaaggag 201

Scoring table:　IDENTITY_NUC
　　　　　　　　　Gapop 10.0 , Gapext 1.0

Searched:　27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters:　55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
　　　　　　　　　 Maximum Match 100%
　　　　　　　　　 Listing first 45 summaries

Database :　EST:*
　　　　1: em_estba:*
　　　　2: em_esthum:*
　　　　3: em_estin:*
　　　　4: em_estmu:*
　　　　5: em_estov:*
　　　　6: em_estpl:*
　　　　7: em_estro:*
　　　　8: em_htc:*
　　　　9: gb_est1:*
　　 10: gb_est2:*
　　 11: gb_htc:*
　　 12: gb_est3:*
　　 13: gb_est4:*
　　 14: gb_est5:*
　　 15: em_estfun:*
　　 16: em_estom:*
　　 17: em_gss_hum:*
　　 18: em_gss_inv:*
　　 19: em_gss_pln:*
　　 20: em_gss_vrt:*
　　 21: em_gss_fun:*
　　 22: em_gss_mam:*
　　 23: em_gss_mus:*
　　 24: em_gss_pro:*
　　 25: em_gss_rod:*
　　 26: em_gss_phg:*
　　 27: em_gss_vrl:*
　　 28: gb_gss1:*
　　 29: gb_gss2:*

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
c 1	79	39.3	644	28	BZ249455	CH230-314
c 2	43	21.4	362	28	BH858390	B5_216a2
3	36.8	18.3	1002	13	BU538230	AGENCOURT
4	35.6	17.7	755	14	CD438041	EL01N0508

No.	Score	%	Length	DB	ID	Description
c 5	35.6	17.7	1068	9	AL527496	AL527496
c 6	35.4	17.6	699	29	CE672514	tigr-gss-
7	35.2	17.5	875	14	CD327143	AGENCOURT
8	35.2	17.5	982	29	CNS0142R	AL103629 Drosophil
c 9	35.2	17.5	1004	29	CNS0066X	AL065920 Drosophil
10	34.2	17.0	1201	9	AL525275	AL525275
c 11	34	16.9	607	14	CB581228	AMGNNUC:N
12	34	16.9	717	10	BE391709	60128236
13	34	16.9	995	28	CC292246	CH261-610
14	33.8	16.8	604	14	CB237820	AGENCOURT
15	33.4	16.6	702	28	AZ847437	2M0148G04
16	33.2	16.5	313	29	AG263957	Lotus cor
17	33.2	16.5	489	13	AZ160863	SP_0067_B
18	33.2	16.5	714	13	BX509001	DKFZp686O
19	33.2	16.5	730	12	BJ014817	BJ014817
20	33.2	16.5	787	13	BU537610	AGENCOURT
21	33	16.4	539	28	AZ092244	RPCI-23-4
22	32.8	16.3	331	9	AI414501	ma39g07.x
23	32.8	16.3	355	14	CD969718	QAC6g06.s
24	32.8	16.3	479	13	BY240517	BY240517
25	32.8	16.3	616	28	AZ396298	1M0160J16
26	32.8	16.3	710	28	CC310470	TAM32-10N
27	32.6	16.2	399	14	R81779	yj03a01.s1
28	32.6	16.2	480	28	AZ329846	1M0054J21
29	32.6	16.2	480	28	AZ515154	1M0054J21
30	32.6	16.2	563	14	CD479241	eca01-33m
31	32.4	16.1	371	9	AA587556	nd44c09.s
32	32.4	16.1	518	28	AQ792699	HS_5299_A
33	32.4	16.1	576	13	BX673778	BX673778
34	32.4	16.1	703	28	CC166880	i15hl1.b
35	32.4	16.1	834	14	CB948166	AGENCOURT
36	32.4	16.1	909	28	CC346613	OGOBA11TH
37	32.4	16.1	1201	9	AL548542	AL548542
38	32.2	16.0	304	10	BB221567	BB221567
39	32.2	16.0	344	29	CE639367	tigr-gss-
40	32.2	16.0	390	14	CB778000	AMGNNUC:T
41	32.2	16.0	435	29	CE543124	tigr-gss-
42	32.2	16.0	462	14	CB606450	AMGNNUC:T
43	32.2	16.0	470	12	BG375690	UI-R-CSO-
44	32.2	16.0	538	14	CB717464	AMGNNUC:N
45	32.2	16.0	563	14	CC609216	AMGNNUC:N

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

ALIGNMENTS

RESULT 1
BZ249455/c
LOCUS   BZ249455   644 bp   DNA   linear   GSS 12-OCT-2002
DEFINITION CH230-314P20.TJ CHORI-230 Segment 2 Rattus norvegicus genomic clone
            CH230-314P20, genomic survey sequence.
ACCESSION   BZ249455
VERSION   BZ249455.1 GI:23910687
KEYWORDS   GSS.
SOURCE   Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1 (bases 1 to 644)
AUTHORS   Zhao,S., Shetty,J., Shatsman,S., Tsegaye,G., Geer,K.,
            Shvartsbeyn,A., Gebregeorgis,E., Overton,L., Russell,D., Chen,D.,
            Riggs,F., de Jong,P. and Fraser,C.M.
TITLE   Rat BAC End Sequences from Library CHORI-230 MboI segment
JOURNAL   Unpublished (1999)
COMMENT   Other_GSSs: CH230-314P20.TV
            Contact: Shaying Zhao
            Department of Eukaryotic Genomics
            The Institute for Genomic Research
            9712 Medical Center Dr., Rockville, MD 20850, USA
            Tel: 301 838 0200
            Fax: 301 838 0208
            Email: szhao@tigr.org

```
         Clones are derived from the rat BAC library CHORI-230
         (http://www.chori.org/bacpac/rat230.htm). For BAC library
         availability, please contact Pieter de Jong (pdejong@mail.cho.org).
         Clones may be purchased from BACPAC Resources
         (http://www.chori.org/bacpac/or ering_information.htm). BAC end
         page: http://www.tigr.org/tdb/bac_ends/rat/bac_end_intro.html
         Plate: 314  row: SP6 column: 20
         Seq primer: SP6
         Class: BAC ends.
FEATURES             Location/Qualifiers
     source          1..644
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /strain="BN/SsNHsd/MCW"
                     /db_xref="taxon:10116"
                     /clone="CH230-314P20"
                     /sex="Female"
                     /cell_type="Brain"
                     /clone_lib="CHORI-230 Segment 2"
                     /note="Vector: pTARBAC1.3; Site 1: MboI; Site 2: MboI;
                     CHORI-230 Rat (BN/SsNHsd/MCW) BAC library produced by
                     Pieter de Jong"
ORIGIN

Query Match         39.3%;  Score 79;  DB 28;  Length 644;
Best Local Similarity   73.9%;  Pred. No. 1.1e-11;
Matches 156;  Conservative   0;  Mismatches  45;  Indels  10;  Gaps   4;

Qy    1 ACAAATGTTCTCCCTGAAGCCCTTCCCTGCCAACCAGACCAGCAACTTCCAAAATTC 60
        |||||||| ||| || |||||| ||||||||||||| ||||| ||||||||||||
Db  325 ACACATGTTTATCTGTCGCCCTTCTCCCTGCCAACCGGACCAACGACTTCCAGAAGGT 266

Qy   61 TGCCCGTG---TTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAA 117
        | |  ||    ||||| |  |  |||| |  ||||| | |||| |||| | | | |
Db  265 TCTACAAGGCATAGAGCCTTTCCGTAGGGACCTCTCAGTGCTTCTGAGGAGCGGACGAG 206

Qy  118 G-GGGATCTGATTA----TTGAGGTGTGGAAGGAATAAATAATCAGTCCACAAATAAACA 172
        | |||||| ||| |    ||||||||| |||||||||||||| |||||||||||||||||
Db  205 CAGGGATCCGATGAAAGACTGGAGGTGTTGAAGGAATAAATAACCAGTCCACAAATAAACA 146

Qy  173 AACTGTCC--GGGATTCCTAGAGGGAAGGAG 201
        |||||||   ||||||||||||||||||||
Db  145 AACTGTCCCGGGGATTCCTAGAGGGAAGGAG 115


RESULT 2
BH858390/c
LOCUS       BH858390             362 bp    DNA     linear   GSS 13-NOV-2002
DEFINITION  B5_216a2_t7 Mouse Retroviral Tagged Cancer Gene Database Mus
            musculus genomic clone B5_216a2, genomic survey sequence.
ACCESSION   BH858390
VERSION     BH858390.1  GI:21709211
KEYWORDS    GSS.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 362)
AUTHORS     Suzuki,T., Shen,H., Akagi,K., Morse,H.C., Malley,J.D., Naiman,D.Q.,
            Jenkins,N.A. and Copeland,N.G.
TITLE       New genes involved in cancer identified by retroviral tagging
JOURNAL     Nat. Genet. 32 (1), 166-174 (2002)
MEDLINE     22194816
PUBMED      12185365
COMMENT     Contact: Copeland NG
            Mouse Cancer Genetics Program
            National Cancer Institute
            Bldg. 539, Rm. 229, Frederick, MD 21702-1201, USA
            Tel: 301 846 1260
            Fax: 301 846 6666
            Email: copeland@ncifcrf.gov
            Class: PCR with specific primers.
FEATURES             Location/Qualifiers
```

```
     source          1..362
                     /organism="Mus musculus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10090"
                     /clone="B5_216a2"
                     /sex="female"
                     /tissue_type="leukemia"
                     /clone_lib="Mouse Retroviral Tagged Cancer Gene Database"
                     /note="Inverse PCR method
                     (http://genome2.ncifcrf.gov/RTCGD)"
ORIGIN

Query Match         21.4%;  Score 43;  DB 28;  Length 362;
Best Local Similarity   67.0%;  Pred. No. 0.15;
Matches 77;  Conservative   0;  Mismatches  35;  Indels   3;  Gaps   1;

Qy    1 ACAAATGTTCTCCCTGAAGCCCTTCCCTGCCAACCAGACCAGCAACTTCCAAAATTC 60
        |||||||| |   || |||||| ||||| ||||||| ||||| |||| ||||||
Db  124 ACAAATGTTTATCTGTCGCCCTTCTCCCTACCCAACCGGACCAACAACTTCCAGAAGGT 65

Qy   61 TGCCCGTG---TTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGG 112
        | | | |    |||||||| ||||||| || | ||| ||| |||||||||| ||
Db   64 TCTCGGAGGCATAGAGCCATTCCGTAGGACCATCTCGGTGCTTCTGAGGAAGCGG 10


RESULT 3
BU538230
LOCUS       BU538230            1002 bp    mRNA    linear   EST 13-SEP-2002
DEFINITION  AGENCOURT 10186385 NIH_MGC_107 Homo sapiens cDNA clone
            IMAGE:6568756 5', mRNA sequence.
ACCESSION   BU538230
VERSION     BU538230.1  GI:22848671
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 1002)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/,
            National Institutes of Health, Mammalian Gene Collection (MGC)
TITLE       Unpublished (1999)
JOURNAL     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
COMMENT     cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            cDNA Sequencing by: Agencourt Bioscience Corporation
            DNA Library distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2754  row: p  column: 04
            High quality sequence stop: 102.
FEATURES             Location/Qualifiers
     source          1..1002
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6568756"
                     /tissue_type="adenocarcinoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_107"
                     /note="Organ: breast; Vector: pOTB7; Site_1: EcoRI;
                     Site_2: XhoI; cDNA made by oligo-dT priming.
                     Directionally cloned into EcoRI/XhoI sites using the
                     following 5' adaptor: GGCACGAG(G). Library constructed by
                     Ling Hong in the laboratory of Gerald M. Rubin (University
                     of California, Berkeley) using ZAP-cDNA synthesis kit
                     (Stratagene) and Superscript II RT (Life Technologies).
                     Note: this is a NIH_MGC Library."
ORIGIN

Query Match         18.3%;  Score 36.8;  DB 13;  Length 1002;
Best Local Similarity   63.6%;  Pred. No. 9;
```

Matches   56; Conservative   0; Mismatches   32; Indels   0; Gaps   0;

```
Qy    8  TTCTCCCGAAGCCCTTCCTGCCAACCAGCAGCAACTTCCAAAATTCTGCCCGT     67
         ||||||||||||| || | |  |  ||||||||| ||||| | |||||
Db  795  TTTTCCCGAAGCCCGTTACGGGGCCCCGGGACCCACCCACTTACAACGGTTTCCCCGT  854

Qy   68  GTTTAGCCTTGTTAAGGGGTGTCTCAC     95
         |||||||||  | | |||||||| |
Db  855  TTTTCAATTTTTACAAGGGGTTTTTAAC    882
```

RESULT 4
CD438041
LOCUS   CD438041
DEFINITION   EL01N0508E07.b Endosperm_5 Zea mays cDNA, mRNA sequence.
ACCESSION   CD438041
VERSION   CD438041.1 GI:31353684
KEYWORDS   EST.
SOURCE   Zea mays
ORGANISM   Zea mays
             Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Zea.
REFERENCE   1 (bases 1 to 755)
AUTHORS   Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,E. and Messing,J.
TITLE   Sequencing of the maize endosperm ESTs
JOURNAL   Unpublished (2002)
COMMENT   Contact: Lai, Jinsheng
             Dr. Joachim Messing's lab
             Waksman Institute, Rutgers University
             190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
             Tel: 732-445-3801
             Fax: 732-445-5735
             Email: jlai@waksman.rutgers.edu
             Seq primer: T3.

FEATURES   Location/Qualifiers
     source 1..755
             /organism="Zea mays"
             /mol_type="mRNA"
             /cultivar="W22"
             /db_xref="taxon:4577"
             /clone_lib="Endosperm_5"
             /tissue_type="Endosperm_5"
             /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2: XhoI"

ORIGIN

Query Match   17.7%; Score 35.6; DB 14; Length 755;
Best Local Similarity   58.5%; Pred. No. 19;
Matches   62; Conservative   0; Mismatches   44; Indels   0; Gaps   0;

```
Qy   91  CTCACTCCTTCAGGGAAAGTGGGAAAAGGGGATCTGATTATTGAGGTGTGAAGGAATAA    150
         |||| || | ||| ||| || |   || || || | || |||| ||| |  | ||| ||
Db  225  CTTACATCACCGGTCAGGTCGGACAGGGAGGAGCAGGTAATACTGGTGTTGCTGTCATGC   284

Qy  151  ATAATCAGTCCACAAATAAACAAACTGTCCCGGATTCCTAGAGGGA    196
         | || ||| ||| ||ATAA CA|| |||||| | | |||| ||||||
Db  285  TTTATGAGCCCAGATATAAGCACAGTGTCAGCATTTTTAAAGGGA     330
```

RESULT 5
AL527496
LOCUS   AL527496   1068 bp   mRNA   linear   EST 23-MAY-2003
DEFINITION   AL527496 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens cDNA clone CS0DC024YG07 5-PRIME, mRNA sequence.
ACCESSION   AL527496
VERSION   AL527496.2 GI:31065347
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE   Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   On Feb 13, 2001 this sequence version replaced gi:12790989.
             Contact: Genoscope
             Genoscope - Centre National de Sequencage
             BP 191 91006 EVRY cedex - France
             Email: seqref@genoscope.cns.fr; Web: www.genoscope.cns.fr
             Library was constructed by Life Technologies, a division of Invitrogen. This sequence belongs to sequence cluster 6496.f For more information about this cluster, see
             http://www.genoscope.cns.fr/
             cgi-bin/cluster.cgi?seq=CS0DC024AD04QP1&cluster=6496.f. Contact : Feng Liang Email : fliang@lifetech.com URL :
             http://fulllength.invitrogen.com/ InVitroGen Corporation 1600 Faraday Avenue Genoscope sequence ID : CS0DC024AD04QP1.

FEATURES   Location/Qualifiers
     source 1..1068
             /organism="Homo sapiens"
             /mol_type="mRNA"
             /db_xref="taxon:9606"
             /clone="CS0DC024YG07"
             /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
             /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
             /note="1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-strand cDNA was digested with Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."

ORIGIN

Query Match   17.7%; Score 35.6; DB 9; Length 1068;
Best Local Similarity   43.4%; Pred. No. 20;
Matches   33; Conservative   24; Mismatches   19; Indels   0; Gaps   0;

```
Qy  103  GGGAAGTGGGAAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAATAATCAGTCCA  162
         || | |||||| ||| :| :|||||||||||| :|: :|||: ::|:::|:: ::::|||
Db  866  GGAAARGGGGRAAAARRAAWTTKWWWWWTKKKKKKKKKKDAAAAAAAAAAAWTWAA     925

Qy  163  CAAATAAACAAACTGT    178
         ||| ||| ||| :::
Db  926  AAAAAAAAMAAAAWKK    941
```

RESULT 6
CB672514/c
LOCUS   CB672514   699 bp   DNA   linear   GSS 29-SEP-2003
DEFINITION   tigr-gss-dog-17000329437427 Dog Library Canis familiaris genomic, genomic survey sequence.
ACCESSION   CB672514
VERSION   CB672514.1 GI:36991514
KEYWORDS   GSS.
SOURCE   Canis familiaris (dog)
ORGANISM   Canis familiaris
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
REFERENCE   1 (bases 1 to 699)
AUTHORS   Kirkness,E.F., Bafna,V., Halpern,A.L., Levy,S., Remington,K., Rusch,D.B., Delcher,A.L., Pop,M., Wang,W., Fraser,C.M. and Venter,J.C.
TITLE   The dog genome: survey sequencing and comparative analysis
JOURNAL   Science 301 (5641), 1898-1903 (2003)
MEDLINE   22875432
PUBMED   14512627
COMMENT   Contact: Kirkness EF
             The Institute for Genomic Research
             Department of Eukaryotic Genomics, TIGR, 9712 Medical Center Drive,
             Rockville, MD 20850, USA
             Tel: 301-838-0200
             Fax: 301-838-0208
             Email: ekirkness@tigr.org
             Class: shotgun.

FEATURES   Location/Qualifiers

```
        source          /organism="Canis familiaris"
                        /mol_type="genomic DNA"
                        /strain="standard Poodle"
                        /db_xref="taxon:9615"
                        /clone_lib="Dog Library"
                        /note="Site_1: BstXI; Libraries were prepared from
                        peripheral blood"
ORIGIN

Query Match          17.6%;   Score 35.4;   DB 29;   Length 699;
Best Local Similarity 59.4%;   Pred. No. 22;
Matches      60;   Conservative  0;   Mismatches  41;   Indels  0;   Gaps  0;

Qy    64 CCGTGTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAAGGGGAT 123
            ||||||  ||||       |||    |||  ||| | |||| ||| ||      || |||||
Db   180 CTGTCTTGGTATAACAAAAGGGTTGATGTGTCTTCGCAGGGAACATAGGAACATAGGGAA 121

Qy   124 CTGATTATTGAGGTGTGGAAGAAATAAATAATCAGTCCACA 164
            |||||||||||  |||     |||||||  |||||  |||||
Db   120 ATGGGAAAGGTGTATGGAGAAACAAAATCCTCATTTACCA 80


RESULT 7
CD327143/c
LOCUS       CD327143          875 bp    mRNA    linear   EST 28-MAY-2003
DEFINITION  AGENCOURT 14146982 NICHD_XGC_Eye1 Xenopus laevis cDNA clone
            IMAGE:6947793 5', mRNA sequence.
ACCESSION   CD327143
VERSION     CD327143.1  GI:31091474
KEYWORDS    EST.
SOURCE      Xenopus laevis (African clawed frog)
  ORGANISM  Xenopus laevis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
            Xenopodinae; Xenopus.
REFERENCE   1 (bases 1 to 875)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement:
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM14561 row: a  column: 08
            High quality sequence start: 11
            High quality sequence stop: 683.
FEATURES             Location/Qualifiers
     source          1..875
                     /organism="Xenopus laevis"
                     /mol_type="mRNA"
                     /db_xref="taxon:8355"
                     /clone="IMAGE:6947793"
                     /dev_stage="adult"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NICHD_XGC_Eye1"
                     /note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 2.3 kb. Constructed by Life
                     Technologies. Note: This is a Xenopus Gene Collection
                     (XGC) library."
ORIGIN

Query Match          17.5%;   Score 35.2;   DB 14;   Length 875;
Best Local Similarity 51.2%;   Pred. No. 25;
```

```
Matches      82;   Conservative  0;   Mismatches  78;   Indels  0;   Gaps  0;

Qy    42 CCAGCAACTTCCAAAATTCTGCCCGTGTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTC 101
            ||||  ||| ||     |||| |  ||||| || |     | ||| ||||| || |
Db   586 CCAGTTCCATGTATCTGCTCCACTGAATGGCGTTGTGAGCAGGGTGAGCCCCTCTGA 527

Qy   102 AGGGAAAGTGGGAAAAGGGGATCTGATTATTGAGGTGTGGAAGAAATAATAATCAGTCC 161
            | ||  | || |   || | |||| ||||    |||||     ||| |||| | ||| ||
Db   526 ACTGGCGTGTTAGACAATAGAGAGAATTAGACTTGTTTCTAACTGATTCATTCATCAT 467

Qy   162 ACAAATAAACAAACTGTCCGGATTCCTAGAGGAGGAAGGAG 201
            ||||| |||||||||||   | |   |  |  || || ||||
Db   466 AAAAAAAAACAAACTGTACAGAAAGACATGACGGCAGGAG 427


RESULT 8
CNS0142R
LOCUS       CNS0142R          982 bp    DNA     linear   GSS 26-JUL-1999
DEFINITION  Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN11A16 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION   AL103629
VERSION     AL103629.1  GI:5615240
KEYWORDS    GSS.
SOURCE      Drosophila melanogaster (fruit fly)
  ORGANISM  Drosophila melanogaster
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
            Ephydroidea; Drosophilidae; Drosophila.
REFERENCE   1 (bases 1 to 982)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     Determination of this BAC-end sequence was carried out as part of a
            collaboration with the European Drosophila Genome Project (EDGP) -
            http://www.edgp.ebi.ac.uk -.This Drosophila melanogaster BAC
            library (Dros BAC) was made by Alain Billaud at CEPH (Centre
            d'Etude du Polymorphisme Humain) with funding provided by a MRC
            project grant. The DNA was prepared from embryos by Alain Bucheton
            and Genevieve Payan. It has been constructed in the vector
            pBeloBAC11.
FEATURES             Location/Qualifiers
     source          1..982
                     /organism="Drosophila melanogaster"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:7227"
                     /clone="BACN11A16"
                     /clone_lib="DrosBAC"
                     /plasmid="pBeloBAC11"
                     /note="end : SP6"
ORIGIN

Query Match          17.5%;   Score 35.2;   DB 29;   Length 982;
Best Local Similarity 33.5%;   Pred. No. 26;
Matches      52;   Conservative 39;   Mismatches  64;   Indels  0;   Gaps  0;

Qy    20 CCCTCTTCCCTGCCAACCAGCAGCAACTTCCAAAATTCTGCCCGTGTTAGCCTTGT 79
            :|||  |::|  :  :|:  :  |:   ||   | |  ||    | :  |  :: |:
Db   828 CCCCCCCCCCCCCCCCCSKGTGKTGSWGATTKKBTGKMGDCTSKHGGWGTNTCRKRTATA 887

Qy    80 TAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAAGGGGATCTGATTATTGAGGTGT 139
            ||  :: |  | : |    :|:  :  : : |  | | |||  ::|: ||  ::::| | :
Db   888 TAKGGRDWTGAARMRRKRGTGGGGSRADGKGKKTCKKGKKGGTGWGTATAKKGKKKKKGK 947

Qy   140 GGAAGGAATAAATAATCAGTCCACAAATAAACAAA 174
            :  ::|| | ||  |  :|:|| ||  |    ||:|
Db   948 KKKKTKWAAAAAADAYAAAWKAAWKATAWAAAWA 982


RESULT 9
CNS006TX/c
```

LOCUS   CNS006TX   1004 bp DNA   linear GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC #
      BACR14N23 of RPCI-98 library from Drosophila melanogaster (fruit
      fly), genomic survey sequence.
ACCESSION AL065920
VERSION   AL065920.1 GI:4944888
KEYWORDS   GSS.
SOURCE   Drosophila melanogaster (fruit fly)
 ORGANISM Drosophila melanogaster
      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
      Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 1004)
AUTHORS   Genoscope.
TITLE   Direct Submission
JOURNAL   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
      BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)
      - Web : www.genoscope.cns.fr)
COMMENT   Determination of this BAC-end sequence was carried out as part of a
      collaboration with the Berkeley Drosophila Genome Project (BDGP).
      The BDGP is constructing a physical map of the Drosophila
      melanogaster genome using these BACs. For further information
      please see http://www.fruitfly.org The BDGP Drosophila
      melanogaster BAC library was prepared by Kazutoyo Osoegawa and
      Aaron Mammoser in Pieter de Jong's laboratory in the Department of
      Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
      NY. The library is named RPCI-98 and was constructed by partial
      EcoRI digestion of Drosophila DNA provided by the BDGP from the
      isogenic strain y2; cn bw sp, the same strain used for the BDGP's
      P1 and EST libraries. A more detailed description of the library
      and how to order individual BAC clones, the entire library, or
      filters for hybridization from the BACPAC Resource Center can be
      found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES   Location/Qualifiers
 source   1..1004
       /organism="Drosophila melanogaster"
       /mol_type="genomic DNA"
       /db_xref="taxon:7227"
       /clone="BACR14N23"
       /clone_lib="RPCI-98"
       /note="end : TET3"
ORIGIN

Query Match   17.5%; Score 35.2; DB 29; Length 1004;
Best Local Similarity 55.6%; Pred. No. 26;
Matches 40; Conservative 11; Mismatches 21; Indels 0; Gaps 0;

Qy 103 GGGAAAGTGGGAAAAGGGGATCTGATTATTGAGGTGTGGAAGGAATAAAATAATCAGTCCA 162
      | | ||||| |||||||| | ::||:| | ::|||:|||||| ||| ||:|||
Db 483 GAGAGAGGGGGAAAAARKAAAWKAAAWKATKTTKAATKKGTGRAAGTRAAAANGAAAAAGAGAAAA 424

Qy 163 CAATAAACAAA 174
      ||| || |||
Db 423 AAAANAAAAAA 412

RESULT 10
AL525275
LOCUS   AL525275   1201 bp mRNA   linear EST 22-MAY-2003
DEFINITION AL525275 Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED Homo sapiens
      cDNA clone CS0DC011YM07 3-PRIME, mRNA sequence.
ACCESSION AL525275
VERSION   AL525275.2 GI:31043530
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
 ORGANISM Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1201)
AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE   Full-length cDNA libraries and normalization
JOURNAL   Unpublished (2001)
COMMENT   On Feb 13, 2001 this sequence version replaced gi:12278768.

Contact: Genoscope
      Genoscope - Centre National de Sequencage
      BP 191 91006 EVRY cedex - France
      Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
      Library was constructed by Life Technologies, a division of
      Invitrogen. This sequence belongs to sequence cluster 4913.f For
      more information about this cluster, see
      http://www.genoscope.cns.fr/
      cgi-bin/cluster.cgi?seq=CS0DC011AG04NP1&cluster=4913.f. Contact :
      Feng Liang Email : fliang@lifetech.com URL :
      http://fulllength.invitrogen.com/ InVitrogen Corporation 1600
      Faraday Avenue Genoscope sequence ID : CS0DC011AG04NP1.
FEATURES   Location/Qualifiers
 source   1..1201
       /organism="Homo sapiens"
       /mol_type="mRNA"
       /db_xref="taxon:9606"
       /clone="CS0DC011YM07"
       /tissue_type="NEUROBLASTOMA COT 25-NORMALIZED"
       /clone_lib="Homo sapiens NEUROBLASTOMA COT 25-NORMALIZED"
       /note="1st strand cDNA was primed with a NotI-oligo(dT)
       primer. Five prime end enriched, double-strand cDNA was
       digested with Not I and cloned into the Not I and EcoR V
       sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN

Query Match   17.0%; Score 34.2; DB 9; Length 1201;
Best Local Similarity 26.7%; Pred. No. 49;
Matches 35; Conservative 49; Mismatches 47; Indels 0; Gaps 0;

Qy 15 TGAAGCCCTCTTCCCTGCCCAACCAGACCAGCAACTTCCAAAATTCTGCCGCGTGTTTAGC 74
     |.: :|||:| ::::: :::::: :::: :|: :|||: ::::|| : :||.|:::
Db 1032 TKTMMCCCCCCMMCCMWMWMMMMWMMMMMMMMARGKKKGGGKDDKTTTTTKKKKTTKGKK 1091

Qy 75 CTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAGTGGGAAAAGGGGATCTGATTATTGA 134
     :|:: :||:| :::| :: |: |||:||| ||||:| :: :|||:
Db 1092 KKKGKKKKGKKGGGGKKKKTTTDTATKGGGGSAAAAAGGSAAAAAGKTKDKATTKGKKTKGKG 1151

Qy 135 GGTGTGGAAGG 145
     ||:| |:|::
Db 1152 GGGGGGRARDK 1162

RESULT 11
CB581228/c
LOCUS   CB581228   607 bp mRNA   linear EST 03-APR-2003
DEFINITION AMGNNUC:NRHY7-00006-B1-A nrhy7 (10850) Rattus norvegicus cDNA clone
      nrhy7-00006-b1 5', mRNA sequence.
ACCESSION CB581228
VERSION   CB581228.1 GI:29525269
KEYWORDS   EST.
SOURCE   Rattus norvegicus (Norway rat)
 ORGANISM Rattus norvegicus
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
      Rattus.
REFERENCE 1 (bases 1 to 607)
AUTHORS   Amgen EST Program.
TITLE   Amgen Rat EST Program
JOURNAL   Unpublished (2003)
COMMENT   Contact: Dan Fitzpatrick
      Amgen, Inc
      One Amgen Center Drive, Thousand Oaks, CA 91320-1799, USA
      Tel: 805 447-4881
      Plate: 00006 row: b column: 1.
FEATURES   Location/Qualifiers
 source   1..607
       /organism="Rattus norvegicus"
       /mol_type="mRNA"
       /db_xref="taxon:10116"
       /clone="nrhy7-00006-b1"
       /clone_lib="nrhy7 (10850)"
       /note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; W Rat

ORIGIN
hypothalamus adult female Wistar rat avg. insert size 2.3
kb fraction 6 and 7"

Query Match   16.9%;   Score 34;   DB 14;   Length 607;
Best Local Similarity   50.0%;   Pred. No. 54;
Matches   85;   Conservative   0;   Mismatches   85;   Indels   0;   Gaps   0;

```
Qy   22 CTCTCCCTGCCCACCAGACCAGACAACTTCCAAAATTCGCCCGTGTTTAGCCTTGTTA  81
Db  599 CCCTTTCCAACTCAGTGAGGGCAAGAATGTCCTTGAAGCAGCCTGAGACAACCACTGTGT 540

Qy   82 AAGGGGTGTCTCACTCCTTCAGGGAAGTGGGAAAGGGGATCTGATTATTGAGGTGTGG  141
Db  539 AGGAGGGTCAACTGCCTCAACGTAAGGAAGCAGGGGTTAACAGAGTACTGACCACAGG   480

Qy  142 AAGGAATAAATAATCAGTCCACAAATAAACAAACTGTCCCGGATTCCTAG           191
Db  479 TTGGGTGACAGGGTATGTGCCTAGGAAGGCAAACGGGTTGGGGGGTACTTG          430
```

RESULT 12
BE391709
LOCUS   BE391709   717 bp   mRNA   linear   EST 21-JUL-2000
DEFINITION   601282336F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3604257 5', mRNA sequence.
ACCESSION   BE391709
VERSION   BE391709.1   GI:9337074
KEYWORDS   EST.
SOURCE   Homo sapiens (human)
ORGANISM   Homo sapiens
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1   (bases 1 to 717)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
    Email: cgapbs-r@mail.nih.gov
    Tissue Procurement: ATCC
    cDNA Library Preparation: Ling Hong/Rubin Laboratory
    cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
    DNA Sequencing by: Incyte Genomics, Inc.
    Clone distribution: MGC clone distribution information can be
    found through the I.M.A.G.E. Consortium/LLNL at:
    http://image.llnl.gov
    Plate: LLCM250   row: g   column: 10
    High quality sequence start: 100
    High quality sequence stop: 687.
FEATURES   Location/Qualifiers
source   1..717
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="IMAGE:3604257"
    /tissue_type="endometrium, adenocarcinoma cell line"
    /lab_host="DH10B (phage-resistant)"
    /clone_lib="NIH MGC 44"
    /note="Organ: uterus; Vector: pOTB7; Site_1: XhoI; Site_2:
    EcoRI; cDNA made by oligo-dT priming. Directionally
    cloned into EcoRI/XhoI sites using the following 5'
    adaptor: GGCACCAG(G). Library constructed by Ling Hong
    in the laboratory of Gerald M. Rubin (University of
    California, Berkeley) using ZAP-cDNA synthesis kit
    (Stratagene) and Superscript II RT (Life Technologies)."
ORIGIN

Query Match   16.9%;   Score 34;   DB 10;   Length 717;
Best Local Similarity   54.9%;   Pred. No. 55;
Matches   67;   Conservative   0;   Mismatches   55;   Indels   0;   Gaps   0;

```
Qy   59 TCTGCCCGTGTTTAGCCTTGTTAAAGGGGTGTCTCACTCCTTCAGGGAAGTGGGAAAAG 118
```

Db 384 TGTGCCCTGATGCTTGTATGTTTATGGTAGCCTTTCAGCTGTTTTGACAAAGCGTGGCAAAG 443
Qy 119 GGGATCTGATTATTGAGGTGTGGAAGGAATAATCAGTCCACAATAAACAAACTGT   178
Db 444 GAGAAGTAACTTTTGAGGATGTTAAGCAAGTTTTGGCAGACCAATTCATCAACATAT   503
Qy 179 CC 180
Db 504 TC 505

RESULT 13
CC292246
LOCUS   CC292246   995 bp   DNA   linear   GSS 13-MAY-2003
DEFINITION   CH261-6105_RM1.1 CH261 Gallus gallus genomic clone CH261-6105,
    genomic survey sequence.
ACCESSION   CC292246
VERSION   CC292246.1   GI:30663687
KEYWORDS   GSS.
SOURCE   Gallus gallus (chicken)
ORGANISM   Gallus gallus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
    Phasianinae; Gallus.
REFERENCE   1   (bases 1 to 995)
AUTHORS   Kremitzki,C., Higginbotham,J., Wylie,K., Carter,J., McPherson,J.,
    Warren,W., Graves,T., Mardis,E. and Wilson,R.
TITLE   Gallus gallus BAC End Reads
JOURNAL   Unpublished (2003)
COMMENT   Contact: Richard K. Wilson
    Washington University School of Medicine
    Genome Sequencing Center
    Email: submissions@watson.wustl.edu
    Insert Length: 182000   Std Error: 0.00
    Seq primer: RM1 TACGACTCACTATAGGGAGA
    Class: BAC ends
    High quality sequence start: 46
    High quality sequence stop: 706.
FEATURES   Location/Qualifiers
source   1..995
    /organism="Gallus gallus"
    /mol_type="genomic DNA"
    /strain="Red Jungle Fowl"
    /db_xref="taxon:9031"
    /clone="CH261-6105"
    /sex="female"
    /cell_line="UCD001, inbred 256"
    /clone_lib="CH261"
    /note="Vector: pTARBAC2.1; Site_1: EcoRI; Site_2: EcoRI;
    CH261 Female Chicken library - For library and clone
    ordering information: http://www.chori.org/bacpac"
ORIGIN

Query Match   16.9%;   Score 34;   DB 28;   Length 995;
Best Local Similarity   52.9%;   Pred. No. 56;
Matches   73;   Conservative   0;   Mismatches   65;   Indels   0;   Gaps   0;

```
Qy   17 AAGCCCTCTTCCCTGCCCAACCAGCACCAGCAACTTCCAAAATTCTGCCCGTGTTTAGCCT  76
Db  779 AGGTCCCTTCCCATGATTGCTCAAACATTAACTTTCCACTTTCCTGCTAACTTTAACCT  838

Qy   77 TGTTAAAGGGGTGTCTCACTCCTTCAGGGAAAGTGGGAAAAGGGGATCTGATTATTGAGG 136
Db  839 TGAAAAATGTGAATTTGTCCTTTCAAGAAATGCAAGAAAATGCAAGAAAAGGGCTTCAAAATTTTAAGA 898

Qy  137 TGTGGAAGAATAAATAA 154
Db  899 AGGGAAAGTTTGGAAAA 916
```

RESULT 14
CB237820
LOCUS   CB237820   604 bp   mRNA   linear   EST 10-FEB-2003

DEFINITION   AGENCOURT 11477053 NIH_MGC_166 Mus musculus cDNA clone
              IMAGE:30252220 5', mRNA sequence.
ACCESSION   CB237820
VERSION   CB237820.1 GI:28289398
KEYWORDS   EST.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 604)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE   National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
              Email: cgapbs-r@mail.nih.gov
              Tissue Procurement: T. Rajendra Kumar and Martin M. Matzuk
              cDNA Library Preparation: CLONTECH Laboratories, Inc.
              cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
              DNA Sequencing by: Agencourt Bioscience Corporation
              Clone distribution: MGC clone distribution information can be
              found through the I.M.A.G.E. Consortium/LLNL at:
              http://image.llnl.gov
              Plate: NDCM18 row: j column: 05
              High quality sequence stop: 194.
FEATURES   Location/Qualifiers
    source   1..604
              /organism="Mus musculus"
              /mol_type="mRNA"
              /db_xref="taxon:10090"
              /clone="IMAGE:30252220"
              /lab_host="DH10B (phage-resistant)"
              /clone_lib="NIH_MGC_166"
              /note="Organ: brain; Vector: pDNR-LIB; Site 1: SfiI
              (ggccattggcc); Site 2: SfiI (gccggctcggcc); and 3' adaptor
              sequence: 5'-CACGGCCATTATGGCC-3' and 3' adaptor sequence:
              5'-ATTCTAGAGAGCCGAGCGCGGCCGACATG-dT(30)BN-3' (where B = A,
              C, or G and N = A, C, G, or T). Average insert size 2.05
              kb (range 1.0-4.0 kb). 15/15 colonies contained inserts
              by PCR.This library was enriched for full-length clones
              and was constructed by Clontech Laboratories (Palo Alto,
              CA.Corp."
ORIGIN
Query Match   16.8%;   Score 33.8;   DB 14;   Length 604;
Best Local Similarity 52.5%;   Pred. No. 62;
Matches   74;   Conservative   0; Mismatches   67;   Indels   0;   Gaps   0;

Qy   42 CCAGCAACTCCAAAATTCTGCCCGTGTTAGCCTTGTTAAAGGGGTGTCTCACTCCCTTC 101
              || ||| ||| ||||| |||||||||| ||||||||||||||
Db   388 CCCACACCTTTATAAAAAAAAAACCCTTTTTATTCTTTTTTAAAGGGGGGGGAAAACAGGG 447

Qy   102 AGGGAAAGTGGGAAAAGGGGATCTGATTATTGAGGTGTGAAGAAATAAATAATCAGTCC 161
              ||| ||||| |||||||| || ||||| |||||||| |||||
Db   448 GGGGTTAAAAAAAAAAGGTTTTTTTAAAAGAAGAAGGGGGGGGGGTATAAACCCCTTTTT 507

Qy   162 ACAATAAACAAACTGTCCGG 182
              |||| ||| |||||| ||||
Db   508 AAAAAAAAATATAGTGGGGGG 528

RESULT 15
AZ847437
LOCUS   AZ847437   702 bp   DNA   linear   GSS 21-FEB-2001
DEFINITION   2M0148G04F Mouse 10kb plasmid UUGCIM library Mus musculus genomic
              clone UUGC2M0148G04 F, genomic survey sequence.
ACCESSION   AZ847437
VERSION   AZ847437.1 GI:13028699
KEYWORDS   GSS.
SOURCE   Mus musculus (house mouse)
ORGANISM   Mus musculus
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
AUTHORS   1 (bases 1 to 702)
              Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
              Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,
              Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von
              Niederhausern,A. and Wright,D., Weiss,R.
TITLE   Mouse whole genome scaffolding with paired end reads from 10kb
              plasmid inserts
JOURNAL   Unpublished (2000)
COMMENT   Contact: Robert B. Weiss
              University of Utah Genome Center
              University of Utah
              Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
              84112. USA
              Tel: 801 585 5606
              Fax: 801 585 7177
              Email: ddunn@genetics.utah.edu
              Insert Length: 10000   Std Error: 0.00
              Plate: 0148 row: G column: 04
              Seq primer: CGTTGTAAAACGACGGCCAGT
              Class: plasmid ends
              High quality sequence stop: 702.
FEATURES   Location/Qualifiers
    source   1..702
              /organism="Mus musculus"
              /mol_type="genomic DNA"
              /strain="C57BL/6J"
              /db_xref="taxon:10090"
              /clone="UUGC2M0148G04"
              /sex="Male"
              /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
              /clone_lib="Mouse 10kb plasmid UUGCIM library"
              /note="Vector: PWD42nv; Purified genomic DNA from M.
              musculus C57BL/6J (male) was obtained from the Jackson
              Laboratory Mouse DNA Resource
              (http://www.jax.org/resources/documents/dnares/). The DNA
              was hydrodynamically sheared by repeated passage through a
              0.005 inch orifice at constant velocity. The sheared DNA
              was blunt end-repaired with T4 DNA polymerase and T4
              polynucleotide kinase. Adaptor oligonucleotides were
              ligated to the blunt ends in high molar excess. The
              adaptored DNA was purified and size-selected for a 9.5 to
              10.5 kb range using preparative agarose gel
              electrophoresis. Vector DNA was prepared from a derivative
              of pWD42 (gi|4732114|gb|AF129072.1|, a copy-number
              inducible derivative of plasmid R1. The vector was ligated
              with adaptors complementary to the insert adaptors and
              purified. The sheared, adaptored mouse DNA was annealed to
              adaptored vector DNA, and transformed into
              chemically-competent E. coli XL10-Gold (Stratagene) cells
              and selected for ampicillin resistance."
ORIGIN
Query Match   16.6%;   Score 33.4;   DB 28;   Length 702;
Best Local Similarity 58.6%;   Pred. No. 81;
Matches   58;   Conservative   0; Mismatches   41;   Indels   0;   Gaps   0;

Qy   103 GGGAAAGTGGGAAAAGGGGATCTGATTATTGAGGTGTGAAGGAATAAATAATCAGTCCA 162
              ||||||||||||||||||||||| || ||| ||||||| || ||||||||||||||
Db   193 GGGGAAAATGGGTAGAGAGGAAGTAATATTCTGATGTGAAATGAAATGAATAATGAAAA 252

Qy   163 CAAATAAACAAACTGTCCGGGATTCCTAGAGGGAAGGAG 201
              |||||||||||||||||||| ||||||||| ||||||
Db   253 AAAAAGAAAAAGTTGTATGAAACAACAAGAAAAAAATGGG 291

Search completed: March 25, 2004, 15:30:23
Job time : 1523.97 secs